Development and application of advanced methods for electronic structure calculations

Schmidt, Per Simmendefeldt

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Ph.D. Thesis
Doctor of Philosophy

DTU Physics
Department of Physics

Development and application of advanced methods for electronic structure calculations

Per Simmendefeldt Schmidt

Kongens Lyngby 2017

DTU

DTU Physics
Department of Physics
Technical University of Denmark

Fysikvej
Building 307
2800 Kongens Lyngby, Denmark
Phone +45 4525 3344
info@fysik.dtu.dk
www.fysik.dtu.dk

Summary

This thesis relates to improvements and applications of beyond-DFT methods for electronic structure calculations that are applied in computational material science. The improvements are of both technical and principal character.

The well-known GW approximation is optimized for accurate calculations of electronic excitations in two-dimensional materials by exploiting exact limits of the screened Coulomb potential. This approach reduces the computational time by an order of magnitude, enabling large scale applications.
The GW method is further improved by including so-called vertex corrections. This turns out to yield ionization potentials and electron affinities that are in better agreement with experiments for both bulk and 2D materials. This newly developed method requires the calculation of an exchange-correlation kernel known from time-dependent DFT. The computational cost of the kernel is negligible compared with the cost of the GW calculation itself, and the kernel even improves the convergence performance.

Literature shows and this thesis confirms that the representation of the individual atomic elements through their PAW setup crucially affects the results of GW calculations. For this reason, part of this thesis relates to developing and applying a new method for constructing so-called norm-conserving PAW setups, that are applicable to GW calculations by using a genetic algorithm. The effect of applying the new setups significantly affects the absolute band positions, both for bulk and 2D materials. The new PAW setups are used for producing most of the results presented in this thesis.

A lack of accurate experimental and theoretical data on adsorption energies, relevant to surface chemistry and catalysis, are identified. The RPA method and beyond, that is known to yield accurate ground state energies, is used to calculate accurate adsorption energies for a wide range of reactions. The results are in good agreement with experimental values, where available. Additionally, a database consisting of 200 highly accurate adsorption energies is constructed to benchmark the accuracy of current DFT functionals and to guide future development of new xc functionals for DFT, especially useful for surface science.
Given the accuracy of existing DFT functionals, they were in turn applied in search for catalysts to be used in electrochemical methanol production from methane. Two different types of surfaces were investigated for this reaction; the (110) surface of rutile transition metal oxides and a fairly new class of two-dimensional materials called MXenes. Promising candidates were found within the MXenes.

Resumé

Denne afhandling vedrører forbedringer og anvendelser af metoder, der går ud over DFT, til beregning af elektroniske strukturer indenfor computer-baseret materialevidenskab. Forbedringerne er af både teknisk og principiel karakter.

Den velkendte GW-metode er optimeret til nøjagtige beregninger af elektron-excitationer i todimensionale materialer ved at udnytte nøjagtige grænser for det screenede Coulomb-potentiale. Denne fremgangsmåde reducerer beregningstiden med en størrelsesorden, hvilket muliggør applikationer i stor skala.

GW-metoden er yderligere forbedret ved at inkludere såkaldte vertex-korrektioner. Dette viser sig at give ioniseringspotentialer og elektronaffiniteter, som er bedre i overensstemmelse med eksperimenter for både bulk og 2D-materialer. Denne nyudviklede metode kræver beregning af en exchange- og korrelationskerne, der er kendt fra tidsafhængig-DFT. Beregningsomkostningerne for kernen er ubetydelige sammenlignet med omkostningerne for GW-beregningen selv og kernen forbedrer endog beregningernes evne til at konvergere.

Litteratur viser og denne afhandling bekræfter, at repræsentationen af de enkelte atomare elementer gennem deres PAW-opsætning har afgørende indflydelse på resultaterne af GW-beregninger. Af denne grund vedrører en del af denne afhandling at udvikle og anvende en ny metode til konstruktion af såkaldte normbevarende PAW-opsætninger, der kan anvendes til GW-beregninger ved hjælp af en genetisk algoritme. Effekten af at anvende de nye opsætninger påvirker væsentligt de absolutte båndpositioner, både for bulk og 2D-materialer. De nye PAW-opsætninger er brugt til at producere de fleste af de resultater, der er præsenteret i denne afhandling.

Manglende nøjagtige eksperimentelle og teoretiske data om adsorptionsenergier, der er relevante for overfladekemi og katalyse, er identificeret. RPA og mere avanceret metoder, der vides at give nøjagtige grundtilstandsenergier, er benyttet til at beregne nøjagtige adsorptionsenergier for en række reaktioner. Resultaterne er i god overensstemmelse med eksperimentelle værdier, hvor disse er til rådighed. Derudover er en database bestående af 200 meget nøjagtige adsorptionsenergier konstrueret som sammenligningsgrundtal for eksisterende DFT funktionaler samt til brug for fremtidige udviklinger af nye funktionaler, til anvendelse inden for overfladevidenskab.
Efter at have etableret nøjagtigheden af eksisterende DFT functionaler, blev de anvendt til at lede efter katalysatorer, der kan bruges i elektrokemisk produktion af methanol fra methan. To forskellige slags overflader var undersøgt til denne reaktion; (110) overfladen af rutile overgangsmetaloxider samt en relativ ny gruppe af to-dimensionelle materialer, kalder MXenes. Lovende kandidater blev fundet blandt

de undersøgte MXenes.

Preface

This thesis is submitted in candidacy for the Ph.D. degree in physics from the Technical University of Denmark. The work was carried out at the Center for Atomic-scale Materials design at the Department of Physics in the period September 2014 to September 2017, under supervision of professor Kristian S. Thygesen and co-supervision of professor Karsten W. Jacobsen. The Ph.D. project was funded by an internal scholarship from DTU.

Kongens Lyngby, September 1, 2017

Per Simmendefeldt Schmidt

Acknowledgements

I am grateful to Kristian for the opportunity to work on a Ph.D. project within the exciting and relevant field of computational material science. Your enthusiasm, physical intuition and our constructive discussions have always left me inspired.

I would like to thank the rest of the staff at CAMD; Karsten W. Jacobsen, Thomas Olsen, Jacob Schiøtz, Mikkel Strange and Ole H. Nielsen for always being helpful and open to discussion. In particular I would like to thank Jens Jørgen Mortensen for his insight and help regarding any issue coming up in the course of this project related to GPAW, Niflheim or programming in general. Thanks to Marianne for the administrative support.

Thanks to Jens Jørgen, Logi, Bo, Terese, Johnny and Simone in particular for proofreading and helpful discussions during the writing phase.

Finally I would like to thank all the people who has made my time at CAMD great. In no particular order, Simone, Filip, Kirsten, Morten, Sten, Korina, Chris, Mohnish, Nicky, Chengjun, Manuel, Martin, Torsten, Suranjan, Anders, Peter, Jacob, Daniele, Douglas, Henrik, Estefania.

Having written both my bachelor, master and now Ph.D. thesis at CAMD, I consider it the place of my academic upbringing and I feel that I have learned a lot and developed greatly as a person - thanks again!

List of Papers

Paper I
A Benchmark database of surface and adsorption energies from many-body perturbation theory
Per S. Schmidt, Thomas Bligaard and Kristian S. Thygesen
In preparation.

Paper II
Electrochemical methanol production from methane
Logi Arnarson, Per S. Schmidt, Kristian S. Thygesen, Ifan E. L. Stephens and Jan Rossmeisl
In preparation.

Paper III
Efficient many-body calculations of two-dimensional materials using exact limits for the screened potential: Band gaps of MoS_2, h-BN and phosphorene
Filip A. Rasmussen, Per S. Schmidt, Kirsten T. Winther and Kristian S. Thygesen
Phys. Rev. B. 94, 155406 (2016).

Paper IV
Beyond the GW method: Simple vertex correction improves band energies of bulk and two-dimensional materials
Per S. Schmidt, Christopher E. Patrick and Kristian S. Thygesen
Submitted.

Contents

CHAPTER 1

Introduction

Materials and their specific properties play an enormous role in our everyday life. New technologies such as light emitting diodes, lasers, scanning probe microscopes, carbon fiber reinforced plastics, Li-ion batteries, soft lithography, solar cells, the exotic metamaterials and many more would not have been possible without major breakthroughs in material science [1].
Many of the scientific problems facing humanity could be solved if materials with the desired properties were available.
One example of materials having unique properties relates to solar cells, that convert energy of sunlight into electricity by the photovoltaic effect. The first solid state solar cell was built in 1883 from selenium, resulting in a device with an efficiency of around 1%[2]. Research in material science has improved this efficiency and single junction silicon devices are now approaching the theoretical Shockley-Queisser limiting efficiency of around 33%[3]. In 2015 a 4-junction GaInP/GaAs/GaInAsP/GaInAs-based solar cell achieved a new laboratory record efficiency of 46.1%[4]. These improvements are crucial in making solar cells feasible for competitive technologies that could become alternatives to present days fossil fuel combustion.
Another example relates to the maximum number of computations per time unit that CPUs in our electronic devices can carry out. That is determined by the number of transistors present in the CPU's processor. Research in material science has continuously been able to add more transistors by reducing the size of the silicon-based transistor. Eventually, however, silicon transistors will have to approach the size of atoms, which poses a fundamental limit. Moore's law would arguably not apply for very long unless non-silicon substitutes can be found. Promising candidate materials include indium-phosphor/indium-gallium-arsenide (InP/InGaAs) heterojunction transistors which take advantage of the intrinsic characteristics of the III-IV semiconductors[5].

It is clear that breakthroughs in material science have had and will continue to have a significant influence on future technologies.

Material science involve studying how the exact structures of materials at the atomic scale, including type and relative positions of the atoms, influence

the properties and performances of materials at the macro scale. Achievements in material science research rely on a close collaboration between experimental and theoretical scientists, however this thesis relates to computational material science where theoretical work simulates material properties. Theoretical simulations have two main advantages: Firstly, it is possible to isolate and study specific physical phenomena otherwise entangled in the experimental measurement. Secondly, theoretical studies are much more cost effective for screening projects where the goal is to find an excellent material for a given application by systematically searching through a large number of candidates. It is simply not feasible to synthesize and investigate thousands of materials experimentally. Such a screening study could suggest the most promising candidates for experimental studies which will allow conclusion as to the actual applicability of the selected candidate materials.

This thesis is involved with development and application of advanced quantum mechanical methods for electronic many-body systems used to predict important properties of specific classes of materials. One of the most widely used methods for modelling complex systems within computational material science is called the density functional theory (DFT). However, this approach has shown to be insufficient for accurate predictions of several properties[6]. To push the field forward, there is a constant need of theoretical improvements of the methods available, both in terms of computational efficiency and accuracy. This thesis specifically focuses on benchmarking and improving beyond-DFT methods applied to two specific areas of material science, namely two-dimensional (2D) materials and surface chemistry.

Atomically thin 2D materials have revolutionized materials science over just the past few years. Starting with the exfoliation of a single layer of graphene experimentally in 2004[7], the class of materials has expanded rapidly to now including insulators, semiconductors, semimetals, metals and superconductors[8]. Due to a large variety in the properties of these materials, envisioned applications span from battery electrodes over catalysis to photovoltaics and electronic devices[9]. The most well studied 2D material is graphene which has shown exceptionally high electrical and thermal conductivity[10] as well as carrier mobility[11]. Monolayer hexagonal boron-nitride (hBN) proved to be an excellent substrate for graphene[12] and due to its large band gap it has been used as a gate dielectric[13] and tunnel barrier in transistors[14]. MoS_2 is another example of a well-studied 2D material which belongs to the class of transition metal dichalcogenides (TMDs). From bulk to the monolayer case, MoS_2 undergoes a transition from an indirect to a direct band gap

semiconductor accompanied by an increase in photoconductivity, absorption and photoluminescence[15, 16]. Together with the high flexibility and transparency, the TMDs are excellent candidates for optoelectronics. In addition to the extraordinary properties of the single-layered materials, a great potential lies in the possibility of stacking different 2D crystal into new layered materials held together by weak van der Waals forces. These are referred to as van der Waals heterostructures (vdWHs) and are likely to achieve new and superior properties[17]. Several hundreds of 2D materials are presently available experimentally and the number of ways of combining them in stacks of 10-20 layers is essentially endless. In order to take advantage of the full potential of vdWHs it is essential to develop *efficient* and *reliable* computational methods that can be applied to the individual layers, at first. In order to accurately model electronic and optical properties of the 2D materials, it is essential to account for many-body effects. This means going beyond the standard framework of DFT and apply more computationally demanding methods. In particular, the reduced dielectric screening arising from the reduced dimensionality must be treated with care[18].

Chapter 4 in this thesis introduces an *efficient* method for calculating excitations in 2D materials without compromising accuracy. This opens up for the possibility of larger scale screening studies. With *efficiency* introduced in the first part of Chapter 4, *reliability* is discussed in the second part. Two different materials joined at an interface is called a heterostructure. The behaviour of a heterostructure can depend crucially on the alignment of the so-called energy levels at the interface(s). This is not only limited to the 2D vdWHs. For example, it was first proposed to use a heterostructure of bulk GaAs between two layers of bulk AlAs in lasers in 1963[19]. This confined carriers so that lasing could occur at room temperature and with lower threshold currents. Energy levels in bulk solids are greatly affected by variations and uncertainties in the surface orientation/termination which makes a comparison between theoretical and experimental results problematic. 2D materials however, offer a unique opportunity for obtaining accurate energy levels due to their well-defined surface structures. Theoretically however, it turns out that present state-of-the-art methods give rather inaccurate predictions of absolute band positions. This problem has been addressed by developing a new method described in Chapter 4 showing improved energy levels without significantly increasing the computational cost.

The theoretical framework used to investigate the challenges outlined above rely on a range of approximations in order to be computationally feasible. As

more advanced methods are developed, there seems to be a tendency to assume that these approximations apply in the same way as they did for the simpler calculations. In Chapter 2 it is discussed how some of these approximations affect the calculations of excitations in both bulk and 2D materials when employing more accurate many-body methods.

In Chapter 3, focus is on surface science and catalysis. A catalyst increases the rate of a chemical reaction whilst not being consumed. Catalysts are therefore applied in many areas of society; production of industrially important chemicals such as ammonia (NH_3)[20], converting biomass to biofuels[21], as catalytic converters in cars converting toxic gases and pollutants in exhaust gas to less toxic ones etc[22]. This means that the benefits of finding more efficient catalysts can reduce production cost, improve efficiency of devices and produce more environmentally friendly technological solutions. In order to determine if a material works well as a catalyst for a specific reaction it could be necessary to know how it binds the relevant reactants. The application of DFT to the description of surface-adsorbate bond strength is ever increasing. The expected accuracy of such a calculation is less established as the experimental data available for comparison is limited[23]. The work presented in Chapter 3 applies beyond-DFT methods to adsorption energy calculations to establish a theoretical state-of-the-art baseline to which the computationally cheaper DFT methods can be compared. The beyond-DFT results open the opportunity for constructing more accurate methods for surface science incorporating the best of both worlds; the computational speed of DFT and beyond-DFT accuracy.
Having established the accuracy of DFT for adsorption energy calculations, the method is applied to the problem of electrochemically producing methanol from methane. This is especially relevant as a vast amount of methane is burned as waste in oil fields every day due to the lack of infrastructure for storage and transportation[24]. A much better fit with the existing infrastructure would be liquid methanol produced by oxidizing methane[25, 26]. This process requires a catalyst material with high selectivity towards methanol production, avoiding the formation of other products such as CO_2. For this application, we have investigated the catalytic properties of the (110) surface of rutile transition metal oxides as well as a fairly new class of two-dimensional materials termed MXenes.
Finally a brief conclusion and an outlook is presented in Chapter 6.

Excluding Chapter 6, each chapter is initiated by an introduction to the theory concepts important for the main results of the chapter. Key points are

highlighted but detailed derivations are kept to a minimum, with references to the relevant literature given instead. In Chapter 3 and 4, the main results of the chapters are presented in the publications attached at the end of the thesis. A brief introduction is given to each publication, summarizing the motivation of the study, the methods applied and the main conclusion. For detailed discussion of the results and conclusions, the reader is referred to the attached publication.

1.1 Outline

The structure of the thesis is summarized the following:

- **Chapter 2** presents an introduction to the electronic structure theory of ab-initio calculations to obtain ground state properties, namely that of density functional theory. Particular emphasis is on the atomic datasets used as the foundation for all the following calculations.
- **Chapter 3** presents the methodology of improving total energy calculations by employing the random phase approximation and beyond, and its application to surface chemistry.
- **Chapter 4** addresses how excited state properties can be more accurately obtained through the GW method, and how these can be further improved through the inclusion of vertex corrections. Additionally, the improved efficiency for GW calculations of 2D materials obtained in this work is discussed.
- **Chapter 5** touches upon the topic of connecting GW theory with DFT through the optimized effective potential and how that can address the problem of self-consistency in RPA calculations.
- **Chapter 6** presents a brief summary and an outlook.

CHAPTER 2

Electronic Structure Theory

The central issue in this work and in material science in general, is the description of physical and chemical properties of matter. Matter is described as composed of electrons and a few hundred different kinds of nuclei interacting via coulombic, electrostatic forces. By nature, the electrons must be described by quantum mechanics, whereas the more massive nuclei can most often be regarded as classical particles. Since the development of quantum mechanics, the mathematical equations describing the physics at the atomic scale have been well known. The main challenge of material science has therefore been a question of developing useful methodologies to reduce the complexity of the quantum mechanical equations, making them solvable for the systems under consideration. Density functional theory (DFT)[27], one of the most successful approaches in this regard, serves as the foundation for the more advanced methods applied and developed further in this work. DFT is therefore briefly summarized in this chapter. The relevant physical concepts and the mathematical equations are presented with focus on the practical implementation in the electronic structure code, GPAW (Grid-based Projector Augmented Wave method)[28, 29], used throughout this work. A newly developed method for constructing a basic ingredient in the PAW implementation, the ”setup”, is presented to address issues regarding norm-conservation and the applicability of said setups in beyond-DFT methods. The most relevant limitations of DFT are highlighted as well, motivating the need for the more advanced, beyond-DFT methods. The following four chapters are devoted to work done optimizing and advancing these methods.

2.1 The Many-Body Problem

Matter is considered a collection of atomic nuclei and electrons interacting via coulombic, electrostatic forces. The nuclei and electrons determine the physical and chemical properties of matter which at equilibrium are governed

by the time independent-Schrödinger equation (TIDSE)[30]:

$$\hat{H}\Psi(\mathbf{r},\mathbf{R}) = \epsilon\Psi(\mathbf{r},\mathbf{R}),$$

where ϵ is the energy eigenvalue to the corresponding many-body wave function, $\Psi(\mathbf{r},\mathbf{R})$ and $\mathbf{r}$ ($\mathbf{R}$) the coordinates of the electrons (nuclei). The general Hamiltonian of such a system can be written as

$$\begin{aligned}\hat{H} &= \hat{T}_{\text{nuclei}} + \hat{T}_{\text{electrons}} + \hat{V}_{\text{nn}} + \hat{V}_{\text{ee}} + \hat{V}_{\text{ext}} \\ &= -\frac{1}{2}\sum_{I=1}^{P}\nabla_I^2 - \frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \frac{1}{2}\sum_{I=1}^{P}\sum_{J\neq I}^{P}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \\ &\quad + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{I=1}^{P}\sum_{i=1}^{N}\frac{Z_I}{\mathbf{R}_I - \mathbf{r}_i},\end{aligned}$$

where $\mathbf{R}_I$ are nuclei coordinates, $\mathbf{r}_i$ electronic coordinates, Z_I nuclear charges and P and N the number of nuclei and electrons respectively. The five terms are in order: kinetic energy of the nuclei, kinetic energy of the electrons, coulomb energy from interacting nuclei, coulomb energy from interacting electrons and coulomb energy from interaction between electrons and nuclei. All the equilibrium physical and chemical properties of the system can be derived, in principle, by solving the time-independent Schrödinger equation, but in practice, a complete analytic solution is only available in a few cases such as the hydrogen atom or the H_2^+ molecule[30]. The main reason is that the two-body nature of the Coulomb interaction makes the equation not separable.

The repulsive electron-electron interaction makes an electron located at $\mathbf{r}$ prevent other electrons from approaching the same location. This means that the probability of finding an electron at $\mathbf{r}$ depends on the location of all other electrons; this is known as correlation. Correlation makes it so that it is in general necessary to deal with $3(P+N)$ coupled degrees of freedom. A few approximations are introduced to make the problem solvable[30].

The first is the adiabatic approximation where it is assumed that there is no mixing of different electronic stationary states due to interaction with the nuclei. The electrons instantaneously follow the motion of the nuclei, while always remaining in the same stationary state of the electronic Hamiltonian[30]. This means that the many-body wave function can be factorized into an electronic and a nuclear part

$$\Psi(\mathbf{r},\mathbf{R},t) = \sum_i \Theta_i(\mathbf{R},t)\Phi_i(\mathbf{r},\mathbf{R}),$$

where $\Theta(\mathbf{R}, t)$ describes the evolution of the nuclei and $\Phi(\mathbf{r}, \mathbf{R})$ are the electronic eigenstates. The Hamiltonian for the electronic system is then

$$\hat{h} = \hat{T}_{\text{electrons}} + \hat{V}_{\text{ext}} + \hat{V}_{\text{ee}}, \tag{2.1}$$

where the external potential, V_{ext}, can include the static potential set up by the nuclei as well as any external applied potential. The corresponding electronic TIDSE is then:

$$\hat{h}\Phi_i(\mathbf{r}, \mathbf{R}) = E_i(\mathbf{R})\Phi_i(\mathbf{r}, \mathbf{R}).$$

The adiabatic approximation breaks down when the energy related to transitions between electronic eigenstates becomes comparable to the energy associated with nuclear vibrations or nuclear rotations. In principle this means that the adiabatic approximation breaks down for metallic systems because the energy gap is zero. Typical temperatures are however much lower than the electronic Fermi temperature so electronic excitations are confined to a narrow region around the Fermi surface. This means that most properties are affected only slightly by neglecting non-adiabatic contributions from these excitations. The usual approach is to begin with the adiabatic description and afterwards include non-adiabatic terms (electron-phonon interactions) afterwards perturbatively when necessary[30]. The nuclear coordinates will be omitted in the following and the electronic many-body wave function written as $\Psi(\mathbf{r})$.

One is often interested in obtaining the ground state energy, E_{gs}, from which many properties can be derived. The ground state energy and wave functions can in general be found by minimizing the expectation value of the electronic Hamiltonian of Eq. (2.1):

$$E_{\text{gs}} = \min_{\Psi} \langle \Psi | \hat{h} | \Psi \rangle \,,$$

where the minimization is over all many-electron wave functions that produce the same electron density and over all electron densities that integrate to the total number of electrons.

Given the many-electron ground state, $|\Psi_0\rangle$, the electronic ground state energy is therefore directly found from

$$\begin{aligned} E_{\text{gs}} &= \langle \Psi_0 | \hat{T}_{\text{electrons}} | \Psi_0 \rangle + \langle \Psi_0 | \hat{V}_{\text{ext}} | \Psi_0 \rangle + \langle \Psi_0 | \hat{V}_{\text{ee}} | \Psi_0 \rangle \\ &= \langle \Psi_0 | \sum_{i=1}^{N} \hat{h}_1(\mathbf{r}_i) | \Psi_0 \rangle + \langle \Psi_0 | \hat{V}_{\text{ee}} | \Psi_0 \rangle \,, \end{aligned}$$

with $\hat{h}_1 = \hat{T}_{\text{electrons}} + \hat{V}_{\text{ext}}$. As such, the Hamiltonian is split up into one- and two-particle operators. One approximation to obtaining the ground state energy relevant to the work in this thesis is the Hartree-Fock approximation. This is briefly outlined in the next section.

2.2 The Hartree-Fock approximation

Within the Hartree-Fock approximation[31] the ground state of the many-electron system is approximated with a single Slater determinant. Letting $\mathbf{r}$ denote position and spin, a Slater determinant for N electrons is generally written as

$$\begin{aligned}\Phi_0(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) &= \frac{1}{\sqrt{N!}}\det\{\psi_1\psi_2...\psi_N\} \\ &= \frac{1}{\sqrt{N!}}\begin{vmatrix} \psi_1(\mathbf{r}_1) & \psi_1(\mathbf{r}_2) & \cdots & \psi_1(\mathbf{r}_N) \\ \psi_2(\mathbf{r}_1) & \psi_2(\mathbf{r}_2) & \cdots & \psi_2(\mathbf{r}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_N(\mathbf{r}_1) & \psi_N(\mathbf{r}_2) & \cdots & \psi_N(\mathbf{r}_N) \end{vmatrix},\end{aligned}$$

where ψ_i are single electron states obtainable from solving the electronic TIDSE for the equivalent non-interacting system, i.e. the system for which $\hat{V}_{ee} = 0$. Because of the properties of determinants, the wave function fulfils the Pauli principle, $\Phi_0(\mathbf{r}_1, \mathbf{r}_2) = -\Phi_0(\mathbf{r}_2, \mathbf{r}_1)$. Additionally, if two electrons occupy the same orbital, then $\psi_i = \psi_j$, making two rows equal and the determinant is zero, preventing electrons of the same spin to occupy the same state.
Within this approximation, the energy contributions from the one-electron operators simply become[30]

$$\begin{aligned}\langle\Phi|\sum_{i=1}^{N}\hat{h}_1(\mathbf{r}_i)|\Phi\rangle &= \sum_{i=1}^{N}\int \psi_i^*(\mathbf{r})\hat{h}_1(\mathbf{r})\psi_i(\mathbf{r})\,\mathrm{d}\mathbf{r} \\ &= -\frac{1}{2}\sum_{i=1}^{N}\int \psi_i^*(\mathbf{r})\nabla_i^2\psi_i(\mathbf{r})\,\mathrm{d}\mathbf{r} + \int n(\mathbf{r})V_{\text{ext}}(\mathbf{r})\,\mathrm{d}\mathbf{r}\end{aligned}$$

and the electron-electron interaction energy becomes

$$\langle\Phi|\hat{V}_{\mathrm{ee}}|\Phi\rangle = \left\langle \Phi \middle| \frac{1}{2}\sum_{i\neq j}^{N}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|}\middle|\Phi\right\rangle = \frac{1}{2}\sum_{i\neq j}^{N}\Bigg(\underbrace{\iint \psi_i^*(\mathbf{r})\psi_j^*(\mathbf{r}')\frac{1}{|\mathbf{r}-\mathbf{r}'|}\psi_i(\mathbf{r})\psi_j(\mathbf{r}')\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'}_{\text{Hartree}} - \underbrace{\iint \psi_i^*(\mathbf{r})\psi_j^*(\mathbf{r}')\frac{1}{|\mathbf{r}-\mathbf{r}'|}\psi_j(\mathbf{r})\psi_i(\mathbf{r}')\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'}_{\text{exchange}}\Bigg). \tag{2.2}$$

The energy associated with the Hartree term is the classical electrostatic interaction energy between two charge distributions $n(\mathbf{r})$ and $n(\mathbf{r}')$. The energy associated with the exchange term represents an extra *stabilizing energy* arising from the antisymmetry of the wave function.

This is the exact expression for the exchange energy (EXX). Since the true many body wave function is not a Slater determinant, but a linear combination of many Slater determinants, an energy contribution will be missing from the Hartree-Fock approximation. The missing term is the correlation energy.

The ground state energy within the Hartree-Fock approximation, $E_{\mathrm{gs}}^{\mathrm{HF}}$ can then be written as

$$E_{\mathrm{gs}}^{\mathrm{HF}} = T + E_{\mathrm{ext}} + E_{\mathrm{Ha}} + E_{\mathrm{EXX}}.$$

As written here, both the Hartree and exchange term exclude the $i = j$ term and $E_{\mathrm{gs}}^{\mathrm{HF}}$ is as such self-interaction free; there does not exist an energy contribution from the electron's interaction with itself. In practice however, the terms are calculated from the densities that sum up all contributions and there will exist a self-interaction term in the Hartree energy that is exactly cancelled by that in EXX. This becomes troublesome if the EXX is replaced by another approximation to the exchange energy. This will become important later in this thesis.

A different approach to ground state energy calculations is taken through DFT, which will only be outlined very briefly in the following section, as that is not the main computational method applied in the later part of this work.

2.3 Density Functional Theory

DFT offers an, in principle, exact way of obtaining the ground state energy of a system[27, 32]. Within DFT, the ground state energy is written as an explicit

functional of the electronic density rather than the wave functions. From a numerical standpoint, this is of huge importance since the density is a real-valued function of three coordinates as opposed to the complex many-body wave function of 3N variables, with N number of electrons.

Although in principle exact, DFT requires the knowledge of an unknown, but universal, functional which in practice is approximated, the exchange-correlation (xc) functional. From an approximation to the xc functional, the ground state electron density, $n(\mathbf{r})$ is obtained from self-consistenct solution of a fictitious one-electron Schrödinger equation[33]

$$\left(-\frac{1}{2}\nabla^2 + v_{\text{ext}}(\mathbf{r}) + v_{\text{Ha}}(\mathbf{r}) + v_{\text{xc}}(\mathbf{r})\right)\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}) \tag{2.3}$$

and the relation between the wave functions and the density:

$$n(\mathbf{r}) = \sum_i^{\text{occ.}} |\psi_i(\mathbf{r})|^2.$$

The electrons feel the effective potential, $v_s(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + v_{\text{Ha}}(\mathbf{r}) + v_{\text{xc}}(\mathbf{r})$. This is constructed so that the resulting density of the fictitious systems is the same as that of the true many-body system.

The ground state energy becomes a functional of the ground state density and reads

$$E[n] = T_{\text{NI}}[n] + \int n(\mathbf{r})v_{\text{ext}}(\mathbf{r})\mathrm{d}\mathbf{r} + \frac{1}{2}\int\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' + E_{\text{xc}}[n]. \tag{2.4}$$

where $T_{\text{NI}}[n]$ is the kinetic energy of the non-interacting system with the same ground state density. The kinetic energy of the interacting system is not the same as for the non-interacting system; there is missing a component accounting for the fact that the true many-body wave function is not a single Slater determinant. This fraction is included in the last term, the xc energy, whose functional derivative yields the xc potential, $v_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}}{\delta n(\mathbf{r})}$.

In practice the xc energy functional must be approximated. The local density approximation (LDA)[27] has been popular in solid state physics and is also relevant to this work.

The local density approximation to $E_{\text{xc}}[n]$ is designed to be exact for the homogeneous electron gas (HEG) meaning that it assumes that the system at a given point in space, $\mathbf{r}$, can be described as a HEG with density $n(\mathbf{r})$. It is therefore expected to perform well for systems with slowly varying densities. LDA is however shown to give acceptable results for many atomic and molecular systems for which the density does not vary slowly[34]. The xc functional

is given by

$$E_{\mathrm{xc}}^{\mathrm{LDA}}[n] = \int n(\mathbf{r})\epsilon_{\mathrm{xc}}^{\mathrm{HEG}}(n(\mathbf{r}))\,\mathrm{d}\mathbf{r}$$

where $\epsilon_{\mathrm{xc}}^{\mathrm{HEG}}(n(\mathbf{r}))$ is the exchange-correlation energy per electron in a HEG. The exchange part can be written explicitly $\epsilon_x^{\mathrm{unif}} = -\frac{3}{4\pi}(3\pi^2 n)^{1/3}$ and a parametric expression for the correlation part is found from Quantum Monte Carlo calculations[35].

LDA is a local functional and does not account well for long-range interactions between systems with non-overlapping densities. LDA suffers from self-interaction errors in the Hartree term and the (surprisingly) good performance comes in part from a systematic underestimation of the correlation energy and an overestimation of the exchange energy resulting in a fortuitous error cancellation. Regarding shortcomings; LDA is known to overestimate bond energies and adsorption energies and systematically underestimate band gaps[30].

Immense efforts have gone into the development of better xc functionals and today there exist hundreds of different types including the generalized gradient approximations (GGA), meta GGAs, (screened) hybrids, GGA+U, and the non-local van der Waals density functionals. The ones relevant for this work will be introduced throughout the thesis as they are used. With a few exceptions, they all contain parameters that have been optimized for a particular type of problem or class of material. Moreover, they rely on fortuitous and poorly understood error cancellations. This limits the generality and predictive power of the standard xc-functionals whose performance can be highly system dependent.

In Section 3.1, the RPA method for obtaining the ground state energy is introduced as an approach to remedy some of these problems.

2.3.1 The band gap problem

In the derivation of Kohn-Sham theory, ϵ_i enter as Lagrange multipliers ensuring orthogonality between the orbitals ϕ_i of the fictitious, non-interacting particles. ϵ_i and ϕ_i are as such simply mathematical tools and the only relevant physical information carried is that the square of the eigenfunctions sum up to the exact ground state density and that, given the exact xc potential, the highest occupied eigenvalue equals the negative of the ionization potential[30].

The band gap of an N-electron system is defined as the difference between the electron affinity $\mathrm{EA} = E_N - E_{N+1} \equiv -\epsilon_{\mathrm{LUMO}}$ and the ionisation potential

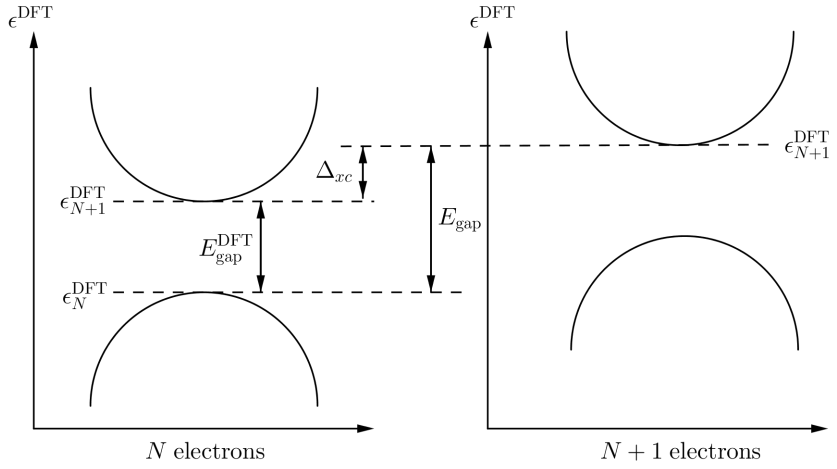


Figure 2.1: Sketch of the DFT band structure of a semiconductor. After addition of an electron occupying the empty conduction band, the xc potential and the band structure shift up by Δ_{xc}.

$\mathrm{IP} = E_{N-1} - E_N \equiv -\epsilon_{\mathrm{HOMO}}$:

$$E_{\mathrm{gap}} = \mathrm{IP} - \mathrm{EA} = \epsilon_{\mathrm{LUMO}} - \epsilon_{\mathrm{HOMO}}$$

where the HOMO (LUMO) are the highest occupied (lowest unoccupied) orbital respectively. Using that the EA of an N-eletron system is the opposite of the IP of the $N+1$ electron system and that the DFT HOMO level is exact, given the exact xc potential, the band gap can be calculated with DFT as:

$$E_{\mathrm{gap}} = \epsilon^{\mathrm{DFT}}_{N+1}[N+1] - \epsilon^{\mathrm{DFT}}_{N}[N]$$

where $\epsilon^{\mathrm{DFT}}_{i}[N]$ is the i-th DFT orbital energy of the N-electron system. This means that $\epsilon^{\mathrm{DFT}}_{N+1}[N+1]$ is the energy of the HOMO level of the $N+1$ electron system and $\epsilon^{\mathrm{DFT}}_{N}[N]$ the HOMO level of the N-particle system.

One usually attempts to calculate the band gap from a single DFT calculation of the N-electron system, resulting in

$$E^{\mathrm{DFT}}_{\mathrm{gap}} = \epsilon^{\mathrm{DFT}}_{N+1}[N] - \epsilon^{\mathrm{DFT}}_{N}[N].$$

This makes the exact and the DFT band gaps related through

$$E_{\mathrm{gap}} = \left(\epsilon^{\mathrm{DFT}}_{N+1}[N] - \epsilon^{\mathrm{DFT}}_{N}[N]\right) + \left(\epsilon^{\mathrm{DFT}}_{N+1}[N+1] - \epsilon^{\mathrm{DFT}}_{N+1}[N]\right) \equiv E^{\mathrm{DFT}}_{\mathrm{gap}} + \Delta_{xc}$$

where we have introduced the derivative discontinuity, Δ_{xc}, which is simply the difference between the energies of the $(N+1)$th orbitals of the N and $N+1$ electron systems[33, 36]. This is sketched in Fig.2.1.

In solids where $N \gg 0$, adding an extra electron will only induce an infinitesimal change of the density. Therefore, the two corresponding KS potentials must be practically the same up to a *constant shift*, and consequently the KS wave functions do not change. Δ_{xc} is this constant shift, entirely contained in v_{xc}, and as such a measure of the non-analytical behaviour of the xc energy functional:

$$\Delta_{xc} = \left(\frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})}\bigg|_{N+1} - \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})}\bigg|_{N} \right) + O\left(\frac{1}{N}\right)$$

If Δ_{xc} were close to zero then the difference between the DFT and the true gap would just be a limitation of the xc functional itself. More sophisticated approaches to the xc energy would then allow us to obtain the real gap directly from its corresponding DFT band-structure. It turns out however that Δ_{xc} is the main cause of the difference between experimental gaps and that from the LDA, which are severely underestimated. It is shown that Δ_{xc} accounts for about 80% of the error.

Δ_{xc} can be calculated by introducing a non-local, energy-dependent self-energy, $\Sigma_{xc}(\mathbf{r}, \mathbf{r}', \epsilon)$, in the Kohn-Sham equation, replacing v_{xc}[36]. Regarding the difference $(\Sigma_{xc} - v_{xc})$ as a perturbation on the Kohn-Sham equation, Δ_{xc} can be approximated from perturbation theory. To first order,

$$\Delta_{xc} = \left\langle \psi_{N+1} \middle| \Big(\Sigma_{xc} - v_{xc} \Big) \Big|_{N-\delta} \middle| \psi_{N+1} \right\rangle,$$

where ψ_{N+1} is the lowest unoccupied Kohn-Sham orbital of the N-electron system and δ a positive infinitesimal.

How to calculate this self-energy and obtain more accurate band gaps from single-electron removal and addition energies with many-body perturbation theory has been a large part of this work and it is presented in Chapter 4. First, however, it is shown how the actual implementation of the DFT formalism can influence the calculated self-energy. This is related to the concept of norm-conserving PAW setups. The implementation of the DFT formalism in the GPAW code is therefore briefly discussed in the following section with emphasis on a newly developed method of constructing so-called norm-conserving setups, crucial to the accuracy of the beyond-DFT method discussed in Chapter 4.

2.4 Implementation of the Kohn-Sham equations in the GPAW code

In addition to the xc potential, more approximations are added in the numerical implementation of solving the Kohn-Sham equations. The Born-Oppenheimer approximation[37] reduced the problem to only involve the electronic wave functions but it is advantageous to further reduce the number of electrons included in the calculations to those that are responsible for most of the physical and chemical properties of a material, the valence electrons. This is in practice done by a frozen core approximation where the core states are assumed to be localized at the individual atoms, unperturbed by the environment, and can as such be calculated once for the isolated atom. How many electrons to "freeze" in the core is a choice to make. Additionally, since the electronic wave functions are eigenstates of the Hamiltonian, they are orthogonal to each other. With the wave functions of the core electrons being localized around the nucleus, the wave functions of the valence electrons must oscillate rapidly in the core region to maintain this orthogonality. This makes them hard to describe accurately without a very large basis set. In the Projector Augmented Wave (PAW) formalism, first described by Blöchl in 1994 [38] and applied in the GPAW code, the rapidly oscillating valence states are mapped onto a basis of smooth computationally convenient pseudo wave functions through a linear transformation inside an augmentation sphere around each atom, defined by a cutoff radius. The number of valence electrons, projector functions, cutoff radii and more determines the "setup" for each element. This will be described in greater detail in the following subsections. The process of generating the smooth pseudo wave functions and the effect they have on the final calculations will be discussed, both for DFT and beyond-DFT methods. The main purpose of this section is to document a newly developed method for using a genetic algorithm[39] to automate the process of constructing setups, a process that has previously taken many man-hours to carry out.

2.4.1 Projector Augmented Wave (PAW) formalism

In the following section the PAW formalism will be presented. As this is not a study of the PAW formalism itself, only the main points relevant to the following sections will be outlined. For details, the reader is referred to [38] or the excellent write-up in [40] and the references therein.

The rapid oscillations of the orthogonal wave functions near the nuclei are

hard to handle numerically. In order to avoid these, we seek a linear transformation which returns the true Kohn-Sham single particle wave function from an auxiliary smooth wave function

$$|\psi_n\rangle = \hat{\mathcal{T}}\,|\tilde{\psi}_n\rangle$$

where n is the quantum state label containing a $\mathbf{k}$, a band and a spin index. This results in the transformed Kohn-Sham equations

$$\begin{aligned}\tilde{\mathcal{T}}^\dagger \hat{H} \tilde{\mathcal{T}}\,|\tilde{\psi}_n\rangle &= \epsilon_n \tilde{\mathcal{T}}^\dagger \tilde{\mathcal{T}}\,|\tilde{\psi}_n\rangle \\ \tilde{\hat{H}}\,|\tilde{\psi}_n\rangle &= \epsilon_n \hat{S}\,|\tilde{\psi}_n\rangle \end{aligned} \tag{2.5}$$

The true wave functions are already smooth at a certain distance from the nuclei so we define $\hat{\mathcal{T}}$ through the atom-centered transformations $\hat{\mathcal{T}}^a$ so that it only modifies the wave functions within an augmentation sphere close to the nuclei defined by a cutoff radius, r_c^a:

$$\hat{\mathcal{T}} = 1 + \sum_a \hat{\mathcal{T}}^a.$$

Inside the augmentation spheres, the true wave function is expanded in partial waves, ϕ_i^a (also called all-electron (AE) partial waves), which in principle has to form a complete set. Each come with a corresponding smooth partial wave (pseudo (PS) partial wave), $\tilde{\phi}_i^a$, which satisfies:

$$|\phi_i^a\rangle = (1 + \hat{\mathcal{T}}^a)\,|\tilde{\phi}_i^a\rangle \Leftrightarrow \hat{\mathcal{T}}^a\,|\tilde{\phi}_i^a\rangle = |\phi_i^a\rangle - |\tilde{\phi}_i^a\rangle$$

As $\hat{\mathcal{T}}^a$ has no effect outside the augmentation sphere, the exact and the smooth partial waves must be identical for $r > r_c^a$.
The smooth partial waves form a complete set inside the augmentation sphere, so we can also expand the smooth all electron wave function as

$$|\tilde{\psi}_n\rangle = \sum_i P_{ni}^a\,|\tilde{\phi}_i^a\rangle\,,$$

where P_{ni}^a are some undetermined expansion coefficients. Applying $\hat{\mathcal{T}}$, one obtains

$$|\psi_n\rangle = \hat{\mathcal{T}}\,|\tilde{\psi}_n\rangle = \sum_i P_{ni}^a\,|\phi_i^a\rangle\,,$$

which shows that the true wave function has identical expansion coefficients, P_{ni}^a. Since $\hat{\mathcal{T}}$ is linear, P_{ni}^a must be linear functionals of $|\tilde{\psi}_n\rangle$:

$$P_{ni}^a = \langle \tilde{p}_i^a | \tilde{\psi}_n \rangle\,,$$

where $|\tilde{p}_i^a\rangle$ are some, to be determined, projector functions. Defining the one center expansion of the smooth wave function as $|\tilde{\psi}_n^a\rangle = \sum_i |\tilde{\phi}_i^a\rangle \langle \tilde{p}_i^a|\tilde{\psi}_n\rangle$, and arguing that it must reduce to $|\tilde{\psi}_n\rangle$ inside the augmentation sphere, the projector functions must satisfy that

$$\sum_i |\tilde{\phi}_i^a\rangle \langle \tilde{p}_i^a| = 1.$$

Hence, the projector functions must be orthonormal to the partial waves inside the augmentation spheres, $\langle \tilde{p}_i^a|\tilde{\phi}_j^a\rangle = \delta_{i,j}$. Given no restrictions outside, we define them to be zero, $\tilde{p}_i^a(r) = 0$ for $r > r_c^a$.

With the completeness relation, we immediately obtain

$$\begin{aligned}\hat{\mathcal{T}}^a &= \sum_i \hat{\mathcal{T}}^a |\tilde{\phi}_i^a\rangle \langle \tilde{p}_i^a| = \sum_i (|\phi_i^a\rangle - |\tilde{\phi}_i^a\rangle) \langle \tilde{p}_i^a| \Rightarrow \\ \hat{\mathcal{T}} &= 1 + \sum_a \sum_i (|\phi_i^a\rangle - |\tilde{\phi}_i^a\rangle) \langle \tilde{p}_i^a|\end{aligned}$$

and the true Kohn-Sham wave function is therefore obtained from the transformation

$$\psi_n(\mathbf{r}) = \tilde{\psi}_n(\mathbf{r}) + \sum_a \sum_i (\phi_i^a(\mathbf{r}) - \tilde{\phi}_i^a(\mathbf{r})) \langle \tilde{p}_i^a|\tilde{\psi}_n\rangle .$$

Subsequently, one usually defines the one-center density matrix

$$D_{i_1,i_2}^a = \sum_n f_n \langle \tilde{\psi}_n|\tilde{p}_{i_1}^a\rangle \langle \tilde{p}_{i_2}^a|\tilde{\psi}_n\rangle .$$

Then it is possible to show that the Kohn-Sham total energy can be separated into a part calculated on smooth functions plus atomic corrections depending on D_{i_1,i_2}^a and system independent tensors that are pre-calculated and stored for each element.

$$\tilde{E}[\tilde{n}] = T_{\mathrm{NI}}[\{\tilde{\psi}_n\}] + E_{\mathrm{Ha}}[\tilde{n}] + E_{xc}[\tilde{n}] + \sum_a \Delta E^a(\{D_{i_1,i_2}^a\})$$

From this, one can obtain the transformed Hamiltonian by carrying out the differentiation

$$\frac{\partial E}{\partial \tilde{\psi}_n^*(\mathbf{r})} = f_n \tilde{H} \tilde{\psi}_n(\mathbf{r}).$$

The Hamiltonian takes the form

$$\tilde{\hat{H}} = -\frac{1}{2}\nabla^2 + u_{\mathrm{Ha}}[\tilde{n}] + v_{xc}[\tilde{n}] + \sum_{a,i_1,i_2} |\tilde{p}_{i_1}^a\rangle \Delta H_{i_1,i_2}^a \langle \tilde{p}_{i_1}^a|$$

where $\tilde{v}_s[\tilde{n}] = u_{\mathrm{Ha}}[\tilde{n}] + v_{xc}[\tilde{n}]$ is the KS potential evaluated on the pseudo density and $\Delta H^a_{i_1,i_2}$ includes the PAW corrections, amongst others a term $\frac{\delta \Delta E_{xc}}{\delta D^a_{i_1,i_2}}$. This becomes important in Chapter 5.

The overlap operator becomes

$$\hat{S} = 1 + \sum_a \sum_{i_1,i_2} |\tilde{p}^a_{i_1}\rangle \left(\langle \phi^a_{i_1} | \phi^a_{i_2}\rangle - \langle \tilde{\phi}^a_{i_1} | \tilde{\phi}^a_{i_2}\rangle\right) \langle \tilde{p}^a_{i_2}|\,,$$

and the all-electron density can be obtained from

$$n(\mathbf{r}) = \sum_n^{\mathrm{val}} f_n |\tilde{\psi}_n|^2 + \sum_a \sum_{i_1 i_2} \left(\phi^a_{i_1}\phi^a_{i_2} - \tilde{\phi}^a_{i_1}\tilde{\phi}^a_{i_2}\right) \sum_n f_n \langle \tilde{\psi}_n | \tilde{p}^a_{i_1}\rangle \langle \tilde{p}^a_{i_1} | \tilde{\psi}_n\rangle + \sum_a \sum_\alpha^{\mathrm{core}} |\phi^{\mathrm{a,core}}_\alpha|^2,$$

The smooth wave function is obtained by solving the transformed Kohn-Sham equation and the a) partial waves, b) smooth partial waves and c) projector functions are system independent and can conveniently be calculated once for each element of the periodic table. a), b), c), a pseudo core density and a local potential define the "setup" for each element. The core density and local potential will not be discussed further but in the following it will be explained how a), b) and c) are calculated in GPAW.

2.5 Setup generation

From the previous section it is evident that PAW calculations require a set of partial waves and projector functions which will be constructed in the order $\phi^a_i \longrightarrow \tilde{\phi}^a_i \longrightarrow \tilde{p}^a_i$. The set must be complete in principle, but this requirement is not fulfilled in practice. A few partial waves per angular momentum channel and per site (= per atom) has proved sufficient. In GPAW the pseudo wave functions can be expanded in plane waves, which is the option relevant to this work.

Omitting spin and shortening the atomic quantum number subscripts $i = (n,l,m)$, $j = (n,l)$, $L = (l,m)$, the partial waves can be written as a product of a radial function times a spherical harmonic

$$\phi_i(\mathbf{r}) = R_j(\mathbf{r}) Y_L(\hat{\mathbf{r}})$$

$Y_L(\mathbf{r})$ are eigenstates of the Laplacian:

$$\nabla^2 Y_L(\theta,\phi) = -\frac{l(l+1)}{r^2} Y_L(\theta,\phi)$$

and $R_j(\mathbf{r})$ is found from solving the radial Schrödinger equation[1]

$$\left[-\frac{1}{2}\frac{d^2}{dr^2} - \frac{1}{r}\frac{d}{dr} + \frac{l(l+1)}{2r^2} + v_s(r)\right] R_j(\epsilon_j, r) = \epsilon_j R_j(\epsilon_j, r), \qquad (2.6)$$

where $v_s(r)$ has to be approximated with LDA or PBE etc. This means that we generate a unique setup for each approximation to the xc potential. The eigenvalue-dependence of the radial wave function is written explicitly.
The radial Schrödinger equation is a second-order linear differential equation and it thus requires two boundary conditions. It can be solved using standard techniques by knowing the asymptotic behaviour of the solution at the origin and at infinity.

Next, in order to construct the smooth partial waves an appropriate augmentation radius, r_c, must be chosen. This should be small enough so that all materials considered subsequently will not have overlapping spheres but large enough so that the core density is well contained within.
Various choices can be made for the mathematical expression for the smooth partial waves; Bessel functions, Gaussians etc but GPAW uses 6th order polynomials of even powers:

$$\tilde{\phi}_i(r) = r^i(a + br^2 + cr^4 + dr^6)$$

where a, b, c, d are determined by matching $\phi_i(r)$ with $\tilde{\phi}_i(r)$ at the cutoff radius

$$\begin{aligned}
\tilde{\phi}_i(r_c) &= \phi_i(r_c) \\
\frac{d}{dr}\tilde{\phi}_i(r)\Big|_{r_c} &= \frac{d}{dr}\phi_i(r)\Big|_{r_c} \\
\frac{d^2}{dr^2}\tilde{\phi}_i(r)\Big|_{r_c} &= \frac{d^2}{dr^2}\phi_i(r)\Big|_{r_c} \\
\frac{d^3}{dr^3}\tilde{\phi}_i(r)\Big|_{r_c} &= \frac{d^3}{dr^3}\phi_i(r)\Big|_{r_c}
\end{aligned}$$

Finally the smooth projector functions, $\tilde{p}_i^a$, are obtained from

$$|\tilde{p}_i^a\rangle = \left(-\frac{1}{2}\nabla^2 + \tilde{v}_s - \epsilon_i\right)|\tilde{\phi}_i^a\rangle$$

The orthogonality condition, $\langle\tilde{p}_i^a|\tilde{\phi}_j^a\rangle = \delta_{i,j}$, is then enforced inside the augmentation spheres by a standard Gram-Schmidt procedure. This specific choice

[1] In practice the scalar relativistic version of the Schrödinger equation is solved.

makes $|\tilde{\phi}_i^a\rangle$ a solution to the transformed Schrödinger equation, Eq.(2.5), meaning that the reference atom is described exactly despite the finite number of projectors.
It might be necessary to include additional partial waves to make the basis sufficiently "complete". The all-electron partial wave is found from solving Eq. (2.6) and the pseudo partial wave from solving the PAW equation

$$\left(\tilde{H} - \epsilon_k \tilde{S}\right)|R_k\rangle = 0 \tag{2.7}$$

for arbitrary ϵ_k and in turn generating the projector function. How well the AE and PS partial waves agree is estimated in terms of the logarithmic derivative at the cutoff radius

$$d^j(R_j) \equiv \frac{1}{R_j(\epsilon_j, r_c)}\left[\frac{dR_j(\epsilon_j, r)}{dr}\right]_{r_c}. \tag{2.8}$$

The following equality for the radial part of the (smooth) partial waves

$$\frac{1}{R_j(\epsilon_j, r_c)}\left[\frac{dR_j(\epsilon_j, r)}{dr}\right]_{r_c} = \frac{1}{R_j(\epsilon_j, r_c)}\left[\frac{d\tilde{R}_j(\epsilon_j, r)}{dr}\right]_{r_c}, \tag{2.9}$$

is strictly fulfilled for energies equal to the eigenvalues, ϵ_j, but not at arbitrary energies; solving the radial Schrödinger equation, Eq.(2.6), to obtain $R_j(\epsilon_j, r)$ or the transformed counterpart, Eq.(2.5), to obtain $\tilde{R}_j(\epsilon_j, r)$ will not necessarily produce wave functions with similar value and derivative ratio. When applying the PAW formalism it is important that the logarithmic derivative agree up to energies relevant for the bands included in the calculations. For DFT calculations, this is most often limited to the valence bands. Agreeing logarithmic derivatives ensure that the smooth partial waves are useful in environments where the eigenvalues do not differ significantly from those used in their construction. E.g. a setup for Si can be used for the bulk solid but also for the Si surface of liquid Si[30].

A term important to setup generation is norm-conservation. This ensures that the partial wave and the smooth counterpart share the same norm inside the cutoff radius

$$\int_0^{r_c} r^2\left[R_j(\epsilon_j, r)\right]^2 dr = \int_0^{r_c} r^2\left[\tilde{R}_j(\epsilon_j, r)\right]^2 dr.$$

There is no fundamental requirement of norm-conservation in the PAW formalism (as is the case for pseudopotentials[30], excluding the ultra-soft pseudopotentials[41]),

only that the partial waves form a complete set inside the cutoff radius. However, if the partial-wave expansion is truncated (as is necessary in practice), the completeness condition is no longer exactly fulfilled. This can become a problem for so-called GW calculations, as presented in Section 2.5.1. The GW method is introduced in great detail in Chapter 4, but for now it is just a beyond-DFT method used to obtain accurate excitation energies. It turns out that re-enforcing norm-conservation can help reducing the error made from not fulfilling the completeness condition. More on this in the following subsection.

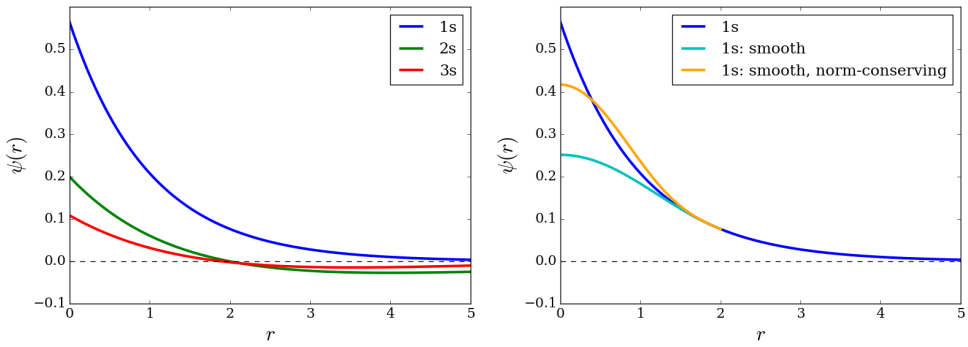
First, a simple example of generating a smooth partial wave is shown for the hydrogen atom. We know the exact wave functions and the s-type is listed below:

- 1s:
$$\psi_{1s}(r) = \frac{1}{\sqrt{\pi a_0^3}} e^{-r/a_0}$$
- 2s:
$$\psi_{2s}(r) = \frac{1}{4\sqrt{2\pi a_0^3}} \left(2 - \frac{r}{a_0}\right) e^{-r/2a_0}$$
- 3s:
$$\psi_{2p}(r) = \frac{1}{81\sqrt{3\pi a_0^3}} \left(27 - \frac{18r}{a_0} + \frac{2r^2}{a_0^2}\right) e^{-r/3a_0}$$

These wave functions are shown in Fig. 2.2a for $a_0 = 1$. In Fig. 2.2b the smooth wave function (6th order polynomial) corresponding to the 1s is shown in cyan with $r_c = 2$. It is immediately clear that the norm is not conserved; $\langle\tilde{\phi}_{1s}|\tilde{\phi}_{1s}\rangle \neq \langle\phi_{1s}|\phi_{1s}\rangle$. If we are to enforce norm conservation, we must simultaneously add a r^8 term to $\phi_l(r)$

$$\tilde{\phi}_l^{\mathrm{nc}}(r) = r^l(a + br^2 + cr^4 + dr^6 + er^8)$$

so that we now solve five equations for five unknown. This yields the wave function in orange in Fig. 2.2b. As a consequence this is less smooth and will thus require a larger basis set of plane waves to be described accurately. This will increase the computational cost so the question is if preserving the norm is important?



(a) (b)

Figure 2.2: (a) Radial part of the 1s, 2s and 3s wave functions of the hydrogen atom. (b) The 1s wave function along with a smooth and a norm-conserving smooth counterpart ($r_c = 2$).

2.5.1 Norm-conservation and GW calculations

The GW method requires evaluation of so-called pair-densities, $\langle n|e^{iG\cdot r}|m\rangle$. Denoting the (smooth) partial wave ($|\tilde{\alpha}\rangle$) $|\alpha\rangle$, the projector function $|p_\alpha\rangle$, the all-electron orbitals $|n\rangle$ are related to the smooth part, $|\tilde{n}\rangle$ by the linear transformation

$$\begin{aligned} |n\rangle &= |\tilde{n}\rangle + \sum_\alpha \Big(|\alpha\rangle - |\tilde{\alpha}\rangle \Big) \langle p_\alpha|\tilde{n}\rangle \\ &= |\tilde{n}\rangle + |n^{\text{aug}}\rangle \,. \end{aligned}$$

The pair-densities will as such involve four terms (here for $G = 0$):

$$\langle n|m\rangle = \langle \tilde{n}|\tilde{m}\rangle + \langle n^{\text{aug}}|m^{\text{aug}}\rangle + \langle n^{\text{aug}}|\tilde{m}\rangle + \langle \tilde{n}|m^{\text{aug}}\rangle \,.$$

With the crucial assumption that the set of partial waves is complete, the pair density simplifies to[38]

$$\langle n|m\rangle = \langle \tilde{n}|\tilde{m}\rangle + \sum_{\alpha\beta} \langle \tilde{n}|\tilde{p}_\alpha\rangle \Big(\langle \alpha|\beta\rangle - \langle \tilde{\alpha}|\tilde{\beta}\rangle \Big) \langle \tilde{p}_\beta|\tilde{m}\rangle \,.$$

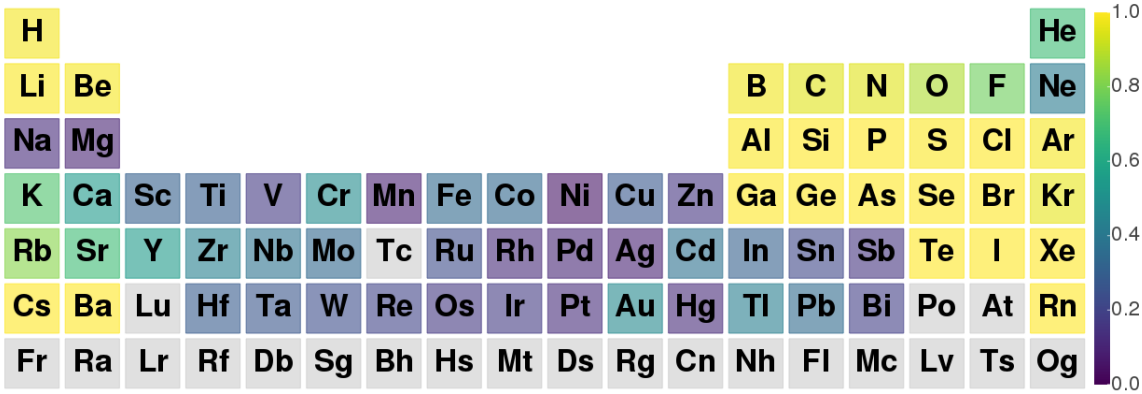
This simplification is crucial to the efficiency of the PAW method. It is however not possible to achieve completeness for orbitals at very high energies, so all four terms should in principle be kept. Take n as a plane wave with large

momentum, then the projection $\langle \tilde{n}|p_\alpha\rangle$ becomes small because the projectors only span the occupied orbitals. That means we are left with

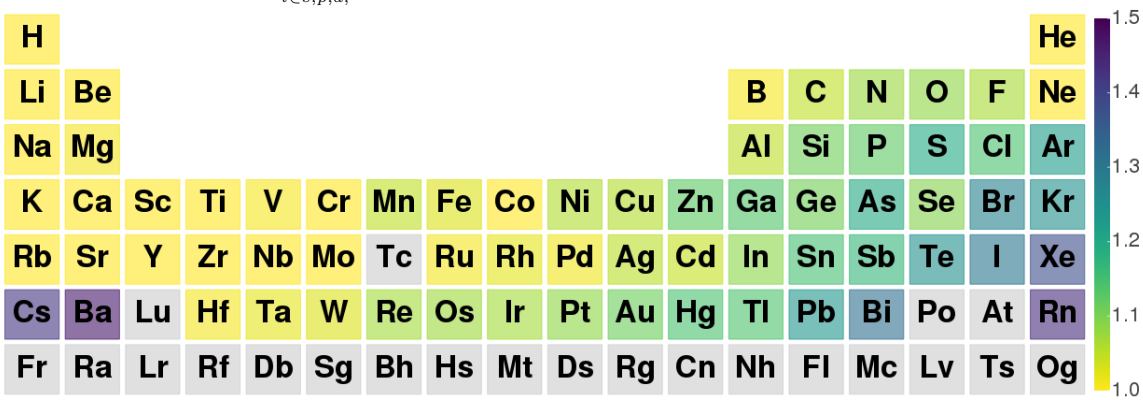
$$\langle n|m\rangle = \langle \tilde{n}|\tilde{m}\rangle \,.$$

Hence if the smooth wave functions have the correct norm, the pair densities will agree exactly for $G = 0$ and approximately for larger plane waves. Norm-conservation should therefore result in a better approximation for the GW self-energy, and it can be seen as a way to reduce the error made by throwing away the cross-terms in the pair densities. The error in the norm can be quite significant for some of the elements with the current way of constructing the setups. In Fig.2.3 the elements are colored according to the (a) minimum and (b) maximum norm of the smooth partial waves. Norm-conservation (norm = 1) is yellow in both figures. The problem is seen to be most present for transition metals, where the norm is violated most significantly for the d electrons.

This problem has been addressed by Klimes et al. in [42] for the VASP code and comparisons to this particular publication will be presented in the following section after introducing how we decided to generate norm-conserving setups for the GPAW code.



(a) $\min_{l \in s,p,d,\cdots} \left(||\tilde{\phi}_l|| \right)$ for the valence electrons of each element.



(b) $\max_{l \in s,p,d,\cdots} \left(||\tilde{\phi}|| \right)$ for the valence electrons of each element.

Figure 2.3: Periodic table colored according to the norm of the smooth partial waves entering the PBE setups of the various elements.

2.5.2 Genetic Algorithm for setup generation

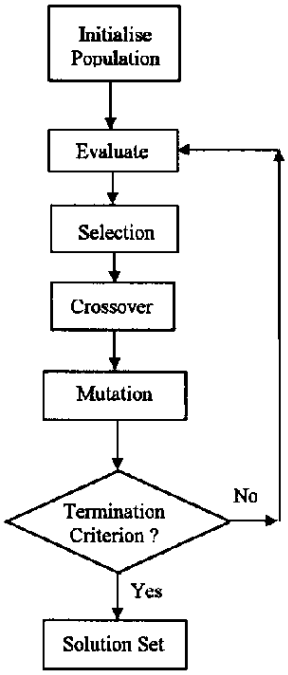


Figure 2.4: Flowchart of a genetic algorithm.

As discussed in the previous section, a PAW setup consist of partial waves, smooth partial waves, projector functions, a local potential and a pseudo core density. Each angular momentum channel of an element will require a cutoff radius for the augmentation sphere. To approach completeness, it might be necessary with an extra partial wave, smooth partial wave and projector function for each angular momentum channel at energies to be determined. Additionally a cutoff for the local potential is needed. This means that around $N = 7$ variables must be determined for each setup. Previously, the N-dimensional search-space has been explored manually which has proven to be a time-consuming and demanding task as simultaneous optimization of accuracy of the setup and computational expense is required. In this regard, evolutionary computing techniques can provide a powerful alternative to finding optimal PAW setups and to eliminate user supervision. A Genetic Algorithm (GA) is one of such techniques which has been applied in this work and elsewhere[43]. We attempted to address the problem of using PAW datasets in beyond-DFT calculations requiring a correct representation of higher-lying unoccupied electronic states. In the following section it is briefly explained how the GA available in the GPAW code works and preliminary results are presented and compared with that of the VASP code which also apply the PAW formalism.

The GA is initiated by creating a population of possible solutions (PAW setups) spanning the N-dimensional search space. A fitness factor is calculated for each setup, which is a number defining how "good" the setup is. What goes into the fitness factor and how to define a good setup will be elaborated upon shortly. The individuals of the population are then evolved through iterations towards optimal fitness under the genetic operations: selection, crossover and mutations. The selection process enforces the "survival of the fittest"-principle, probabilistically favouring individuals with better fitness in becoming parents

of the next generation. Crossover and mutation causes small random and unbiased changes to the individuals in the population. More specifically, crossover searches for better individuals by mixing already good ones and mutation can introduce features that were lost in the current population[43]. This procedure is sketched in Fig.2.4. The GA used here introduces more aggressive crossover and mutation during the first iterations in order to sample widely in the N dimensional search space before narrowing down the search. Currently no termination criteria is set on the fitness factor as the absolute value of the fitness factor turned out to be very element specific. The algorithm now simply runs for the set amount of time.

The fitness factor determining the quality of a setup was chosen to test the following properties:

- Logarithmic derivative error from -1.5 to 2.0 Hartree. Tolerance: 0.3.

$$\text{Log-Error} = \sum_{l\in\{s,p,d\}} \int_{-1.5}^{2.0} \mathrm{d}E \left| d^l(\tilde{R}^l(E)) - d^l(R^l(E)) \right|$$

- Deviation from FHI-AIMS all-electron calculations of fcc and rocksalt (element combined with oxygen) equillibrium lattice constants. Tolerance: 0.01 Å.
- Difference in energy change when compressing the fcc and rocksalt structures with 10 and 20% compared with FHI-AIMS equillibrium value. Tolerance: 0.03, 0.05 eV.
- Number of iterations of the Kohn-Sham equations before convergence of a 7-layer, fcc(111) terminated slab. Tolerance: 40.
- Cutoff required for the total energy of an atom in a $3 \times 3 \times 3$ Å box to be within 0.1 eV of the fully converged calculation (at 1500 eV). Tolerance: 400 eV.
- Eggbox error of the atom in a similar sized box. Eggbox error defined as the largest difference in energy as the atom is placed in four different positions through the unit cell. This is only an issue for the finite-difference implementation in the GPAW code. Tolerance: 0.0005 eV.
- It is punished if the cutoff radii for the different angular momentum channels become larger than the covalent radius of the atom. Tolerance: 0.4 Å.

And the resulting fitness factor calculated as:

$$\text{Fitness factor} = \sum_{\text{properties}} \left(\frac{\text{Error}}{\text{Tolerance}} \right)^2$$

As such, the magnitude of the tolerance determines how much weight a given property is given in the fitness factor.
The search-space is defined by the user when constructing the initial population in terms of how many electrons that are defined to be frozen in the core and how many extra partial waves are included. Additionally, the user decides if norm-conservation is enforced. From there the GA optimizes the cutoff radii of the s, p, d, ... channels and the local potential as well as the energies of the additional partial waves. The tolerance values defined (rather arbitrarily) above will of course play a large role in which setups the GA ends up with. This can for example be used to construct setups which converge quickly at the expense of accuracy which could be useful for first-time quick'n'dirty calculations providing a rough idea of the properties of the system of interest. They could also be used for educational purposes where students are trying to become familiar with the program but not interested in highly accurate output numbers. On the other hand, extremely accurate but computationally slower setups could be constructed for the specific cases where this is required.

The GA has been applied to a range of elements, some of which are of particular importance to this work. The main interest in this regard is how norm-conservation affects the setup when applied in GW calculations.
First, the GA was applied to the simpler elements, namely C, B and N where the wave functions are close to being norm-conserving already and the enforced norm-conservation is not expected to have an effect on the following GW calculations. The performance of the two best C and the best B and N setups in relation to the fitness factor is reported in Table 2.1. It should be noted that the parameters have been discretized so the minimum variation in a parameter across different setups is 0.05 Åand 0.10 eV. Both for the standard case as well as enforced norm-conservation. The GA ran for 50 hours, evaluating almost 2000 different setups for each element. We immediately see that the best and second best C setup are almost identical, and so are the next ten as well (not reported here). This indicates that the GA has located an optimum in the search space. The cutoff energy necessary to achieve convergence is clearly higher for the norm-conservation setups, in good agreement with the fact that the norm-conserving wave-functions are less smooth, requiring more rapidly oscillating plane waves to be described accurately.

Performance:								
	C #1	C #2	B	N	C-nc #1	C-nc #2	B-nc	N-nc
Fitness factor	5.7	6.0	1.4	10.8	5.4	5.4	2.1	19.5
Log. deriv.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
fcc - a_0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.01
fcc - 10%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
fcc - 20%	0.02	0.02	0.03	0.04	0.01	0.00	0.01	0.07
rocksalt - a_0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
rocksalt - 10%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
rocksalt - 20%	0.07	0.07	0.02	0.12	0.03	0.04	0.01	0.15
Iterations	30	33	24	25	24	22	21	31
Cutoff	381	381	250	398	420	440	483	656
Eggbox	0.0004	0.0004	0.0001	0.0008	0.0004	0.0005	0.0002	0.0006
Radii	0.5	0.5	0.1	0.3	0.7	0.6	0.1	0.7
Parameters:								
Radii: s	1.65	1.65	1.50	1.30	1.40	1.35	1.40	1.20
Radii: p	1.30	1.30	1.25	1.35	1.70	1.70	1.25	1.40
Radii: d	1.10	1.10	1.00	1.15	1.30	1.30	1.50	1.60
Proj. energies: s	-0.20	-0.20	0.90	-1.10	1.60	1.30	2.20	-0.30
Proj. energies: p	0.20	0.20	0.80	0.80	-0.20	-0.10	0.70	1.80
Proj. energies: d	1.10	1.20	-0.70	-0.90	2.10	2.20	-2.00	1.10
Radii: Local potential	1.10	1.10	1.00	1.15	0.90	0.90	1.15	1.10

Table 2.1: Parameters for the C, B and N setups discussed in this section.

In Table 2.2, various energy differences through the Brillouin zone from the resulting GW calculations using the setups above as well as the standard (std) setup available with GPAW are reported.

For C we see that the standard setup agrees really well with the one generated using the GA. Additionally there is almost no difference comparing to the setup where norm-conservation has been enforced. We do see deviations comparing with the results from [42] or [44] up to 0.18 eV.
For BN the gap spans 0.15 eV between the three GPAW setups. In GW calculations we worry about converging the results to within 0.02 eV with respect to the number of k-points, plane wave cutoff etc. but this suggests that an even larger error might be introduced in the used setup already. This issue needs to be addressed in the future. The relevant question here is however, how do we determine the correct number for the band gap from a GW calculation? Do we need to compare with all-electron codes? The implementation of D.

	Material	ΔIP	Γ_c	L_v	L_c	χ_v	χ_c	Gap
D. Nabok et al. [44]	C	-0.99	7.43	-2.94	10.38	-6.58	6.26	
J. Klimes et al. [42]	C	-1.07	7.43	-2.92	10.37	-6.55	6.21	
GPAW std	C	-1.02	7.57	-2.99	10.54	-6.62	6.32	5.74
GPAW GA	C	-1.02	7.57	-2.99	10.54	-6.63	6.34	5.75
GPAW GA nc	C	-1.05	7.58	-2.99	10.55	-6.63	6.37	5.78
D. Nabok et al. [44]	BN	-1.47	11.28	-2.04	12.32	-5.13	6.47	6.47
J. Klimes et al. [42]	BN	-1.35	11.29	-2.08	12.25	-5.19	6.37	6.37
GPAW std	BN	-1.42	11.48	-2.07	12.43	-5.17	6.54	6.54
GPAW GA	BN	-1.31	11.40	-2.10	12.37	-5.22	6.45	6.45
GPAW GA nc	BN	-1.47	11.53	-2.07	12.52	-5.15	6.60	6.60

Table 2.2: G_0W_0@LDA results for C and BN: the absolute shift of VBM compared to LDA (ΔIP) and the positions of VBM at χ (χ_v) and L (L_v) and CBM at Γ (Γ_c), χ (χ_c) and L ($_c$) relative to VBM at Γ. Hence Γ_c is the direct band gap. The values are extrapolated to infinite cutoff.

Nabok et al. [44] is in an all-electron code but they compared to the results of J. Klimes et al. [42] to tune their basis. We obviously cannot compare with experimental results as GW is not exact and we should not construct setups which bias GW results towards the experimental values if that does not represent the actual accuracy of the method. Additionally, adding GW calculations to the fitness factor when constructing setups will be way too time consuming on the GA. The GW test must be a post-process so it is important to determine which parameters that go into the setup construction affects the subsequent GW calculations. One such parameter could be the logarithmic derivatives, as this represents how well higher lying states will be described. This will be discussed in greater detail in the following section for an element where these considerations are much more crucial than for B and N, namely Zinc.

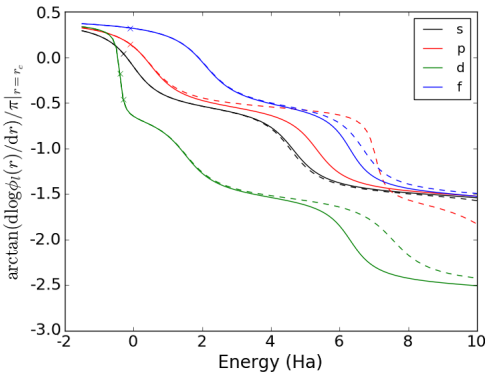
2.5.2.1 Zinc

In [42] it was argued that Zinc was one of the elements where norm-conservation was most severely violated. That is also evident from Fig.2.3. The norm of the 3d partial wave is 0.154. A set of Zinc setups was therefore constructed using the GA algorithm in order to compare their performance with the standard setup and to reproduce the results in [42]. The standard setup for Zinc includes the 3d, 4s and 4p partial waves (12 valence electrons) and includes an extra s, p and d partial wave along with their smooth counterpart and projector functions. The setups generated with the GA also includes the 3s and 3p as valence electrons, as is the case in [42]. Additionally we varied the number of extra partial waves included; only one or two extra d, one s, p and d or one s, one p and two d. Setups were generated both with and without norm-conservation enforced. Deviations in the logarithmic derivatives up to 10 Hartree were punished, ensuring better agreement for higher energies. The performance of the resulting setups for the GA and the parameters necessary to generate them are listed in Table 2.4.

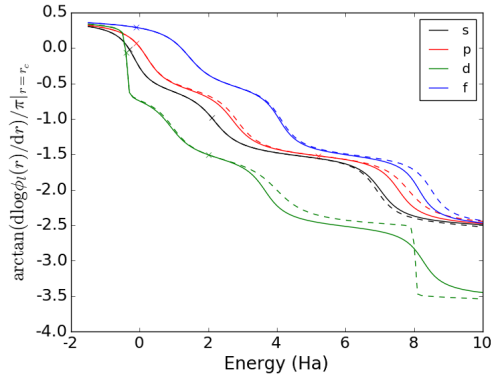
Performance:								
					Norm-conserving			
	Zn d	Zn spd	Zn 2d	Zn sp2d	Zn d	Zn spd	Zn 2d	Zn sp2d
Fitness factor	954.5	192.4	307.5	95.9	51340.6	19796.5	4587.7	862.1
Log. deriv.	1.7	1.3	1.5	0.5	21.7	12.6	4.5	2.3
fcc - a_0	-0.12	0.03	0.01	-0.01	-0.03	0.09	-0.06	-0.01
fcc - 10%	0.30	-0.01	-0.01	0.01	0.66	0.19	0.41	0.08
fcc - 20%	0.03	0.06	0.01	0.00	0.86	0.84	0.20	0.23
rocksalt - a_0	-0.10	0.03	0.01	-0.01	0.07	0.09	-0.21	0.01
rocksalt - 10%	0.31	-0.03	0.02	0.08	-0.35	0.69	0.03	0.03
rocksalt - 20%	0.71	0.02	0.00	0.10	0.23	2.22	0.51	0.09
Iterations	40	57	90	169	56	117	93	106
Cutoff	894	895	754	807	898	889	896	897
Eggbox	0.0063	0.0052	0.0152	0.0125	0.1134	0.0575	0.0843	0.0330
Radii	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Parameters:								
Radii: s	1.15	1.30	1.20	1.30	1.40	1.70	1.40	1.90
Radii: p	1.90	1.15	1.85	1.95	1.85	2.35	1.35	1.60
Radii: d	1.10	1.30	1.85	1.55	1.50	1.90	1.85	1.75
Energies: s	-	4.3	-	3.1	-	2.20	-	2.10
Energies: p	-	6.0	-	5.9	-	5.30	-	5.80
Energies: d	0.6	0.5	1.9, 5.8	-1.3, 6.5	-0.2	2.1	0.9, 5.9	0.5, 5.5
Radii: Local pot.	0.95	1.00	5.80	0.90	1.35	1.20	1.25	1.20

Table 2.3: Performance and parameters for the GA generated Zn setups with various number of extra projectors. The 3s, 3p, 3d, 4s, 4p were treated as valence electrons.

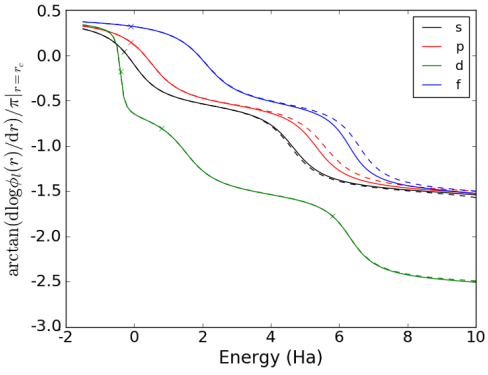
The inverse tangent of the corresponding logarithmic derivatives are shown in Fig.2.5 from -1.5 to 10.0 Hartree, shifted down by π at divergences in the logarithmic derivatives. The dashed line correspond to the smooth partial waves. The matching clearly improves when adding more projectors but this will simultaneously increases the computational cost. We want the logarithmic derivatives to match to as high energies as possible, but it is observed that this becomes more difficult when enforcing norm-conservation. Adding more projectors improve the logarithmic derivatives as well as the other parameters included in the fitness factor. The fitness factor is in general much higher for the norm-conserving setups, mainly due to larger logarithmic derivatives, larger deviations for lattice constants and a larger eggbox error. The only nc setup that shows an acceptable performance is sp2d.



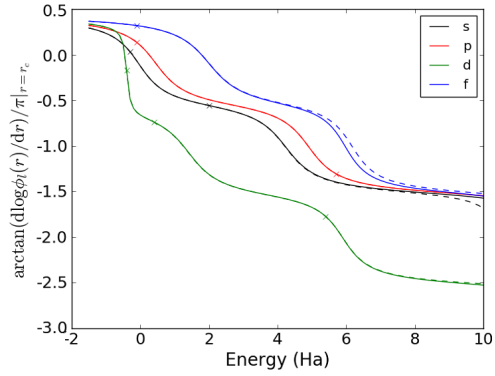
(a) Extra projectors: d



(b) Extra projectors: s,p,d



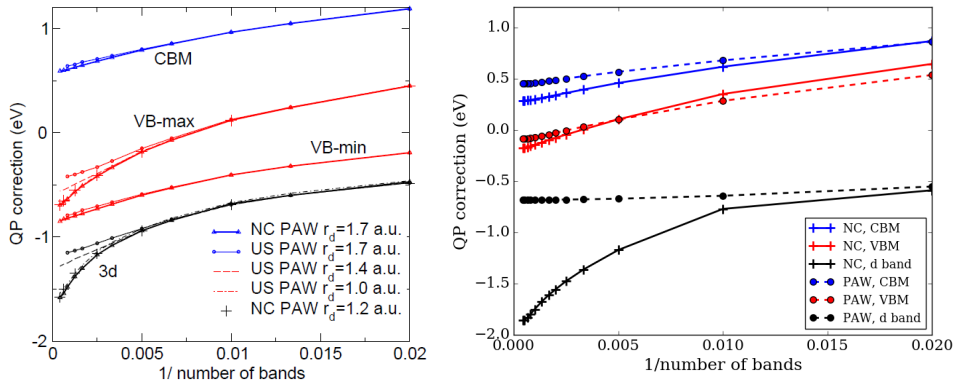
(c) Extra projectors: 2d



(d) Extra projectors: s,p,2d

Figure 2.5: Logarithmic derivatives from -1.5 to 10.0 Hartree for norm-conserving Zn setups. Dashes line corresponds to the smooth partial waves. Crosses at the eigenvalues, ϵ_j.

The norm-conserving Zn sp2d setup is subsequently used to reproduce the particular behaviour of specific bands in ZnO as a function of the number of unoccupied bands included in the calculation of the response function, as presented in [42]. The comparison is shown in Fig.2.6 with (a) the results with VASP and (b) the results of this work. The QP correction ($\epsilon^{\mathrm{DFT}} - \epsilon^{\mathrm{GW}}$) is shown for the CBM, VBM and 3d states at the Γ-point. The calculation is carried out at $4 \times 4 \times 4$ k-points and 800 eV plane wave cutoff. An upward shift of 1 eV is added to the VBM and 3d states. The agreement between the two codes is fairly good and the importance of norm-conservation, especially for the *d* states is clear. The converged *d*-band correction differ by around 1 eV compared with the standard setup available, although both potentials are practically indistinguishable for ground-state calculations. As mentioned in [42], the VBM is dominated by oxygen 2p states but at the Γ-point the Zn-3d states hybridize with the O-2p states forming a bonding and antibonding linear combination. The antibonding linear combination has a strong Zn-3d character and hence the VBM follows the 3d states[42]. The shift in the VBM is however less than observed with VASP. We observe a similar shift in the CBM and hence the band gap remains almost unaffected, where as VASP reports an opening of the gap.



(a) (b)

Figure 2.6: QP corrections to the VBM (maximum), CBM and d-band of ZnO with (a) VASP[42] and (b) GPAW.

QP energies for various points throughout the Brillouin zone has been reported independently in [42] and [44] so it is natural to compare GPAW

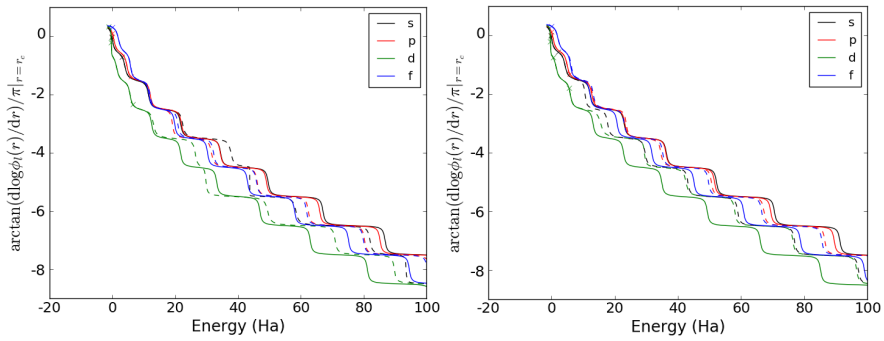
here. The **k**-point sampling is increased and the QP energies are calculated for ZnO in zincblende structure employing the experimental lattice constant and extrapolating to infinite plane wave cutoff including 2200 bands with the 3 different setups generated as well as the standard GPAW setup. The results are printed in Tab.2.4. ΔIP is the absolute shift of VBM compared to LDA and the remaining columns represent the positions of VBM at χ (χ_v) and L (L_v) and CBM at Γ (Γ_c), χ (χ_c) and L ($_c$) relative to VBM at Γ. Hence Γ_c is the direct band gap. None of this is concerned with the 3d states and the numbers (except for ΔIP) are differences between QPs. In Fig.2.6 we saw that the gap at Γ was less affected when employing a norm-conserving setup due to a similar downshift of both the CBM and VBM. The energy differences in column 4 to 8 in Tab.2.4 are seen to be unaffected in a similar way. The deviations from the results of [42] vary from 0.04 to 0.21 eV and from [44] from 0.10 to 0.45 for the norm-conserving setup. The deviations between [42] and [44] vary from 0.01 to 0.31 eV. The fact that codes deviate with up to 0.45 eV for advanced, high-accuracy methods is unsatisfactory and it justifies the need for further investigates as to why. That is however beyond the scope of this work. ΔIP is the same QP correction as plotted in Fig. 2.6 (although for a converged **k**-point sampling in the table) for the VBM. In the figure however, 1.0 eV is added to improve the visualization.

	ΔIP	Γ_c	L_v	L_c	χ_v	χ_c	Gap
D. Nabok et al. [44]	-1.74	2.73	-0.78	7.78	-2.12	7.31	2.73
J. Klimes et al. [42]	-1.53	2.46	-0.79	7.43	-2.15	7.00	2.46
GPAW std	-1.19	2.31	-0.85	7.44	-2.27	7.00	2.31
GPAW GA Zn3s3p spd	-1.20	2.31	-0.85	7.45	-2.27	7.02	2.31
GPAW GA Zn3s3p sp2d	-1.26	2.39	-0.85	7.54	-2.25	7.06	2.39
GPAW GA nc Zn3s3p sp2d	-1.32	2.28	-0.88	7.37	-2.36	7.04	2.28

Table 2.4: G_0W_0@LDA results for ZnO employing various setups: the absolute shift of VBM compared to LDA (ΔIP) and the positions of VBM at χ (χ_v) and L (L_v) and CBM at Γ (Γ_c), χ (χ_c) and L ($_c$) relative to VBM at Γ. Hence Γ_c is the direct band gap. The values are extrapolated to infinite cutoff from calculations at 700 and 800 eV including 2200 bands.

As the fully converged calculations required 800 eV plane wave cutoff, one could argue that having logarithmic derivatives agreeing up to 10 Hartree is not sufficient to accurately capture states up to 800 eV. In Fig.2.7 the logarithic derivatives of the Zn sp2d setups are shown up to 100 Ha. There is a clear deviation for energies above 15 Ha. Possibly not visible from the figure, but the

agreement is worse for the norm-conserving setup. The difference integrated from -1.5 to 100 Hartree summed up for all four angular momenta is 94.2 Ha for the Zn sp2d setup and 132.4 Ha for the norm-conserving one.



(a) (b)

Figure 2.7: Logarithmic derivatives from -1.5 to 100.0 Hartree for (a) the Zn sp2d setup and (b) the norm-conserving Zn sp2d setup.

2.5.2.2 Norm-conserving setups for 2D materials

To carry out GW calculations for the popular 2D materials, WS_2, MoS_2 and $MoSe_2$, norm-conserving setups were generated for Mo and W. The performance and parameters can be found in Appendix A.1. In Fig. 2.8 the VBM and CBM are shown relative to vacuum calculated with PBE (a GGA xc functional)[45], non-selfconsistent HSE06 (hybrid DFT xc functional)[46] and G_0W_0 for $MoSe_2$ in its PBE relaxed structure. Norm-conservation only affects the G_0W_0 calculations: the bands shift down by around 0.5 eV, while the gap closes by 0.09 eV. The valence and conduction bands consist of mostly molybdenum d orbitals, so this case is closely related to that of ZnO, although the shift is not as dramatic as was the case for the d-band of ZnO. A similar picture is observed for WS_2 and MoS_2. This downshift emphasizes the need for norm-conserving setups if one is to obtain the true GW value for the absolute band positions. These can become important for interfaces in heterostructures, for example.

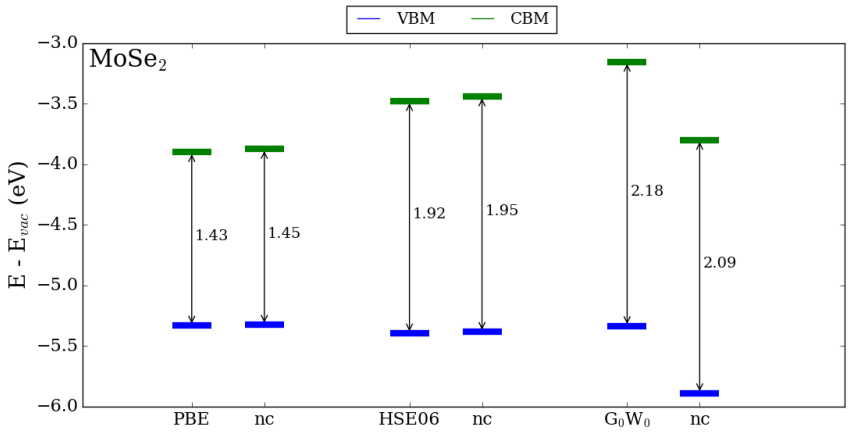


Figure 2.8: Position of VBM and CBM relative to vacuum for 2D $MoSe_2$ calculated with PBE, HSE and G_0W_0 employing the standard and a norm-conserving (nc) setup constructed with the GA for this work.

CHAPTER 3

Improving total energies: The Random Phase Approximation

DFT has proven to be a computationally fast and efficient method for calculating ground state properties for a large variety of systems showing (perhaps surprisingly) accurate predictions of structures and thermodynamic properties of molecules and solids. Despite its success in many areas, it does suffer from large errors causing qualitative failures in predicted properties. Some shortcomings of DFT are the underestimation of barriers of chemical reactions, energies of dissociating molecular ions and charge transfer excitation energies. Furthermore, DFT fails in describing degenerate or near-degenerate states in transition metal systems, breaking of chemical bonds and strongly correlated materials[6]. Most failures are not a breakdown of the theory itself but a deficiency in the xc functional applied. However, a systematic construction of functionals has proven to be an extremely difficult task.

Due to these limitations, efforts have been initiated to find more a fundamental expression for the xc energy. In the following chapter, a formally exact expression for the xc energy will be derived using the adiabatic connection fluctuation-dissipation theorem. The expression requires the knowledge of how the system responds to an external perturbation, i.e. the density response function. This in turn is calculated within the random phase approximation (RPA) which is currently being considered the gold standard for solid state systems due to its ab-initio nature, good description of static correlation and excellent description of long range dispersive forces, such as van der Waals forces[47, 48, 49, 50]. It is then discussed how to go beyond RPA by including a xc kernel from time-dependent DFT and both RPA and beyond methods are in turn applied to the area of surface chemistry.

3.1 Derivation of the Random Phase Approximation

RPA total energy calculations include the exact exchange energy from the Hartree-Fock approximation, Eq. (2.2), in the expression for the total energy

$$E_{\mathrm{xc}}[n] = E_{\mathrm{x}}^{\mathrm{EXX}}[n] + E_{\mathrm{c}}^{\mathrm{RPA}}[n].$$

and the method is concerned with approximating the correlation energy. The RPA expression for the correlation energy can be formulated within different theoretical frameworks, but one convenient approach is the adiabatic connection technique coupled to the fluctuation-dissipation theorem in the DFT context (ACFDT)[49].

Introducing the true many-body wave function $|\Psi\rangle$ and the Kohn-Sham ground state (Slater determinant of non-interacting wave functions), $|\Phi\rangle$, it is evident that the total Hartree + exchange + correlation energy can be written as

$$E_{\mathrm{Hxc}} = \langle\Psi|\hat{T} + \hat{V}_{\mathrm{ee}}|\Psi\rangle - \langle\Phi|\hat{T}|\Phi\rangle\,.$$

Now it is possible to introduce a fictitious system governed by a λ-parameter $\in [0,1]$ that adjusts the interaction between the electrons. The Hamiltonian of the λ-system is

$$\hat{H}(\lambda) = \hat{T} + \lambda\hat{V}_{\mathrm{ee}} + \hat{V}_{\mathrm{ext}}^{\lambda}.$$

While turning on the electron-electron interaction by increasing λ, the external potential is adjusted to keep the same ground state density.
Ψ_λ is now defined as the wave function that minimizes the expectation value of $\hat{H}(\lambda)$ such that $\Psi_{\lambda=0} = \Phi$ and $\Psi_{\lambda=1} = \Psi$. With this, E_{Hxc} can be rewritten

$$\begin{aligned} E_{\mathrm{Hxc}} &= \langle\Psi_\lambda|\hat{T} + \lambda\hat{V}_{\mathrm{ee}}|\Psi_\lambda\rangle\Big|_{\lambda=1} - \langle\Psi_\lambda|\hat{T} + \lambda\hat{V}_{\mathrm{ee}}|\Psi_\lambda\rangle\Big|_{\lambda=0} \\ &= \int_0^1 \mathrm{d}\lambda \frac{\mathrm{d}}{\mathrm{d}\lambda}\langle\Psi_\lambda|\hat{T} + \lambda\hat{V}_{\mathrm{ee}}|\Psi_\lambda\rangle \\ &= \int_0^1 \mathrm{d}\lambda\,\langle\Psi_\lambda|\hat{V}_{\mathrm{ee}}|\Psi_\lambda\rangle\,, \end{aligned} \tag{3.1}$$

where it is used that $\frac{\mathrm{d}}{\mathrm{d}\lambda}\langle\Psi_\lambda|\hat{Q}|\Psi_\lambda\rangle = \langle\Psi_\lambda|\frac{\mathrm{d}}{\mathrm{d}\lambda}\hat{Q}|\Psi_\lambda\rangle$ by the variational principle[33]. The coulomb interaction in second quantization reads

$$\hat{V}_{\mathrm{ee}} = \frac{1}{2}\int\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'\frac{1}{|\mathbf{r}-\mathbf{r}'|}\hat{\psi}^\dagger(\mathbf{r})\hat{\psi}^\dagger(\mathbf{r}')\hat{\psi}(\mathbf{r}')\hat{\psi}(\mathbf{r}),$$

and using the anti-commutator relation $\{\psi(\mathbf{r}), \psi^\dagger(\mathbf{r}')\} = \delta(\mathbf{r}-\mathbf{r}')$ on the middle pair of operators, it can be rewritten into

$$\hat{V}_{\mathrm{ee}} = -\frac{1}{2}\int\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \frac{1}{|\mathbf{r}-\mathbf{r}'|}\left[\hat{n}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}') - \hat{n}(\mathbf{r})\hat{n}(\mathbf{r}')\right],$$

where $\hat{n}(\mathbf{r}) = \hat{\psi}^\dagger(\mathbf{r})\hat{\psi}(\mathbf{r})$. Inserting this into Eq. (3.1) and subtracting the Hartree and exchange energy, $\langle\Psi_{\lambda=0}|\hat{V}_{\mathrm{ee}}|\Psi_{\lambda=0}\rangle$, an expression for the correlation energy is derived

$$\begin{aligned} E_{\mathrm{c}} &= \frac{1}{2}\int_0^1 \mathrm{d}\lambda \int\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \frac{1}{|\mathbf{r}-\mathbf{r}'|}\Big[\Big(\langle\Psi_0|\hat{n}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}')|\Psi_0\rangle - \langle\Psi_\lambda|\hat{n}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}')|\Psi_\lambda\rangle\Big) \\ &+ \Big(\langle\Psi_\lambda|\hat{n}(\mathbf{r})\hat{n}(\mathbf{r}')|\Psi_\lambda\rangle - \langle\Psi_0|\hat{n}(\mathbf{r})\hat{n}(\mathbf{r}')|\Psi_0\rangle\Big)\Big] \\ &= \frac{1}{2}\int_0^1 \mathrm{d}\lambda \int\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \frac{1}{|\mathbf{r}-\mathbf{r}'|}\Big[\langle\hat{n}(\mathbf{r})\hat{n}(\mathbf{r}')\rangle_\lambda - \langle\hat{n}(\mathbf{r})\hat{n}(\mathbf{r}')\rangle_0\Big], \end{aligned} \tag{3.2}$$

where it is used that the expectation value of the density for both the non-interaction and λ-interacting wave functions are the same per construction.

$\langle\hat{n}(\mathbf{r})\hat{n}(\mathbf{r}')\rangle$ is in general called a two-body density matrix and is related to an object called the pair correlation function, $g(\mathbf{r},\mathbf{r}')$ which, physically speaking, represents the probability of finding an electron at $\mathbf{r}$ given that there is one at $\mathbf{r}$'. Because of correlation, the presence of an electron discourages other electrons from approaching it and therefore the pair correlation function must go from 0 at $\mathbf{r} = \mathbf{r}'$ to 1 for $\mathbf{r}-\mathbf{r}' \to \infty$. The relation connecting the two-body density matrix to the pair correlation function is $\langle\hat{n}(\mathbf{r})\hat{n}(\mathbf{r}')\rangle = \frac{1}{2}\langle\hat{n}(\mathbf{r})\rangle\langle\hat{n}(\mathbf{r}')\rangle g(\mathbf{r},\mathbf{r}')$, but finding an explicit expression for pair correlation function, is a major problem in many-body theory[30].

In order to evaluate $\langle\hat{n}(\mathbf{r})\hat{n}(\mathbf{r}')\rangle_\lambda$ it is necessary to introduce an object called the density-density response function. Assuming a system with density $n(\mathbf{r})$ at time t_0 and an external perturbation $V(\mathbf{r},t)$, the density at time $t > t_0$ can be written

$$n(\mathbf{r},t) = n(\mathbf{r},t_0) + \int \mathrm{d}\mathbf{r}' \int_{t_0}^{\infty} \mathrm{d}t' \frac{\delta n(\mathbf{r},t_0)}{\delta V(\mathbf{r}',t')}\delta V(\mathbf{r}',t') + \cdots$$

and the retarded linear density-density response function is then defined as

$$\chi^R(\mathbf{r},\mathbf{r}',t,t') = \frac{\delta n(\mathbf{r},t)}{\delta V(\mathbf{r}',t')}.$$

For a weak perturbation, the response of a system is dominated by the linear effects. The response function of an interacting system is a complicated

object and needs to be approximated, but an analytic expression for the non-interacting response function can be derived from time-dependent perturbation theory[51]

$$\chi_0^R(\mathbf{r},\mathbf{r}',i\omega) = \sum_{i,j} \frac{f_i - f_j}{i\omega + \epsilon_i - \epsilon_j} \psi_i^*(\mathbf{r})\psi_j(\mathbf{r})\psi_i(\mathbf{r}')\psi_j^*(\mathbf{r}'). \tag{3.3}$$

with $f_{i,j}$ being the occupation numbers (0 or 1) and $\epsilon_{i,j}$ the eigenenergies of the corresponding i,j eigenstates, $\psi_{i,j}$. The R for retarded will be omitted from here on. This is a rather computationally expensive object to evaluate since the sum over i and j are over all occupied as well as unoccupied states. This sum is in principle infinite but has to be truncated at a chosen number of unoccupied states determined by the so-called cutoff energy, E_{cut}.

From the Fluctuation-Dissipation theorem it is possible to obtain an expression for $\langle \hat{n}(\mathbf{r})\hat{n}(\mathbf{r}')\rangle_\lambda$ in terms of the response function of the $\lambda-$system. Some details are explained in [49] and the result repeated here:

$$\langle \hat{n}(\mathbf{r})\hat{n}(\mathbf{r}')\rangle_\lambda = n(\mathbf{r})n(\mathbf{r}') - \frac{1}{\pi}\int_0^\infty \mathrm{d}\omega \chi_\lambda(\mathbf{r},\mathbf{r}',i\omega).$$

This theorem relates a density fluctuation to the dissipation of a system represented by the density-density response function on imaginary frequencies.

Inserting this expression into Eq. (3.2) for both the λ-system and the $\lambda = 0$ system gives a final (and still exact) expression for the correlation energy:

$$E_{\text{c}} = -\int_0^1 \mathrm{d}\lambda \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \int\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \left[\frac{1}{|\mathbf{r}-\mathbf{r}'|} \left(\chi_\lambda(\mathbf{r}',\mathbf{r},i\omega) - \chi_0(\mathbf{r}',\mathbf{r},i\omega) \right) \right].$$

The correlation energy can be calculated once an approximation for χ_λ has been made.

From time-dependent DFT, χ_λ can be shown satisfy the Dyson equation[51]

$$\begin{aligned}\chi_\lambda(\mathbf{r},\mathbf{r}',i\omega) = \chi_0(\mathbf{r},\mathbf{r}',i\omega) \\ + \int\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \chi_0(\mathbf{r},\mathbf{r}_1,i\omega) \left(\lambda V_{\text{ee}}(\mathbf{r}_1-\mathbf{r}_2) + f_{\text{xc}}^\lambda(\mathbf{r}_1,\mathbf{r}_2,i\omega) \right) \chi_\lambda(\mathbf{r}_2,\mathbf{r}',i\omega),\end{aligned}$$

where the exchange-correlation kernel is the functional derivative of the xc potential at coupling strength λ, $f_{\text{xc}}^\lambda(i\omega) = \frac{\delta V_{\text{xc}}[n]}{\delta n(\mathbf{r})}$. All correlation effects in χ_λ have been transferred to f_{xc}^λ which, in RPA, is approximated in the simplest possible way as $f_{\text{xc}}^\lambda = 0$ [52]. Various successful attempts at going beyond RPA

($f_{\mathrm{xc}}^{\lambda} \neq 0$) have been made over the recent years. One is particular relevant to this work, known as the renormalized adiabatic LDA (rALDA)[53, 54, 55, 56], and it will be discussed in Section 3.2.

It is common to Fourier transform from real space to **q**-space, which gives $\chi_\lambda(\mathbf{q}, i\omega)$ within RPA as:

$$\chi_\lambda^{\mathrm{RPA}}(\mathbf{q}, i\omega) = \frac{\chi_0(\mathbf{q}, i\omega)}{1 - \lambda V_{\mathrm{ee}}(\mathbf{q})\chi_0(\mathbf{q}, i\omega)},$$

where $V_{\mathrm{ee}}(\mathbf{q}) = 4\pi/|\mathbf{q}|^2$.

The λ-integration can be carried out analytically, resulting in the expression for the RPA correlation energy

$$E_c[n] = \frac{1}{2}\int \frac{d\omega}{2\pi} \mathrm{Tr}\Big\{ \ln[1 - V_{\mathrm{ee}}\chi_0(i\omega)] + V_{\mathrm{ee}}\chi_0(i\omega) \Big\} \tag{3.4}$$

with the trace (Tr) being another (common) way of writing, in real space,

$$\mathrm{Tr}[AB] = \int\int A(\mathbf{r}, \mathbf{r}')B(\mathbf{r}', \mathbf{r})\, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}'$$

or in reciprocal space[57]

$$\mathrm{Tr}[AB] = \frac{1}{N_q} \sum_{\mathbf{q}\in BZ} \sum_{|\mathbf{G}+\mathbf{q}|,|\mathbf{G}'+\mathbf{q}|} A_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) B_{\mathbf{G}\mathbf{G}'}(\mathbf{q}),$$

with N_q being the number of **q**-points, **G** and **G**' reciprocal lattice vectors and BZ short for the first Brillouin zone.

Some advantages of calculating E_{xc} using RPA is that self-interaction errors from the Hartree term is cancelled by that of the exact exchange term. The method does not rely on error cancellation between exchange and correlation energies and it is a non-local method in both the exchange and correlation energy calculations making it possible to account for long range interactions such as the van der Waals interactions. Van der Waals interactions are weak and long ranged compared to e.g. covalent bonds. This makes them rather unimportant in systems with atoms held together by strong covalent bonds, but there does exist systems where vdW interaction is the dominant bonding mechanism. These are e.g. bonding between graphene layers in graphite and inter-molecular forces in a liquid like water[58].

Disadvantages of the RPA include a self-correlation error from the correlation energy term. This is e.g. seen for the hydrogen atom as RPA evaluates a correlation energy for the single electron of approximately -0.5 eV[55]. This

is zero if the exact response function was used. Within the RPA method, an orbital dependent energy functional has to be evaluated and that requires the inclusion of a large number of unoccupied bands, making the method computationally much heavier than standard DFT calculations. Additionally, the RPA correlation energy converges very slowly with the number of unoccupied bands. Luckily, the high energy part of the response function resembles a Lindhard function which gives the following asymptotic limit for the correlation energy[59]

$$E_{\mathrm{c}}^{\mathrm{RPA}} = E_{\mathrm{c}}^{\infty} + \frac{A}{E_{\mathrm{cut}}^{3/2}}. \tag{3.5}$$

From two calculations with different cutoff energies (different amount of unoccupied bands included in the summation of χ_0), it is possible to fit to the equation above and thereby determine A and E_{c}^{∞}. E_{c}^{∞} is then the extrapolated RPA correlation energy at an infinite cutoff energy.

To obtain the ground state energy in RPA, first the ground state wave functions are calculated using the DFT scheme presented earlier within a given approximation to V_{xc}, e.g. LDA. The EXX and RPA correlation energy functionals are then evaluated using the resulting ground state Kohn-Sham wave functions

$$E_{\mathrm{gs}}^{\mathrm{RPA}} = E_{\mathrm{gs}}^{\mathrm{LDA}} - E_{\mathrm{xc}}^{\mathrm{LDA}} + E_{\mathrm{EXX}}[n_{\mathrm{LDA}}] + E_c[n_{\mathrm{LDA}}].$$

This method is often referred to as (EXX+RPA)@LDA and has improved upon pure LDA (or PBE) calculations for a wide range of material properties. One apparent issue is how the RPA results differ depending on which orbitals they are evaluated on (LDA, GGA, ...). This issue is discussed further in Chapter 5. In [57] lattice constants are calculated using RPA and the resulting mean absolute relative error (MARE) compared to experiments is lower for RPA (0.4%) than for PBE (1.4%). Results for RPA calculations of bulk moduli from [57] show that RPA comes out on top with a MARE of 4%, beating both PBE (11%) and LDA (10%). In [60] RPA is shown to predict adsorption energies in good agreement with experiments as well as recover the correct adsorption site for CO on Cu(111). The RPA has however only been applied to a few adsorption reactions. This is a problem which will be discussed in the following section.

Not all properties are improved by RPA compared to the more simple LDA or GGA calculations, as for instance is the case for atomization or cohesive

energies. In [57] RPA atomization energies are shown to have a MARE of 7% where PBE performs better with a MARE of 5 %.

RPA also tends to overestimate absolute correlation energies, as shown in [61]. They differ by more than 30% from the exact correlation energy[57] and in [61] RPA comes out with a mean absolute error of 5.33 eV. This error is attributed to the fact that RPA does not include the so-called short-range second order exchange contribution [57], but it is known to describe the long-range dependence of the correlation energy better. One way of improving the description of short-range correlation relevant to this work is discussed in Section 3.2.

Despite some of the discussed shortcomings, the RPA method is greatly appreciated, the reason being expressed in this quote by J. Harl and G. Kresse (2009) in [60]:

> "*RPA seamlessly incorporates van der Waals interactions, but also describes ionic bonding, covalent and metallic bonding, including d metal bonding and it is universally applicable to solids, molecules and biological systems*",

3.2 Beyond RPA: xc kernels

In this section it is discussed how to go beyond the RPA by the inclusion of a xc kernel, and what the effect is on total energy calculations as discussed in [53, 54, 55, 56]. The xc kernel is applied again in Chapter 4 to improve calculations of excitations.

From the ACFD formulation of the correlation energy we arrived at the formally exact expression:

$$E_{\mathrm{c}} = -\int_0^1 \mathrm{d}\lambda \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \int\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \left[\frac{1}{|\mathbf{r}-\mathbf{r}'|} \left(\chi_\lambda(\mathbf{r}', \mathbf{r}, i\omega) - \chi_0(\mathbf{r}', \mathbf{r}, i\omega) \right) \right], \tag{3.6}$$

which requires the calculation of the response function of the λ-scaled system, χ_λ, from its Dyson equation[51]

$$\begin{aligned} \chi_\lambda(\mathbf{r}, \mathbf{r}', i\omega) = \chi_0(\mathbf{r}, \mathbf{r}', i\omega) \\ + \int\int \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \chi_0(\mathbf{r}, \mathbf{r}_1, i\omega) \left(\lambda V_{\mathrm{ee}}(\mathbf{r}_1 - \mathbf{r}_2) + f_{\mathrm{xc}}^\lambda(\mathbf{r}_1, \mathbf{r}_2, i\omega) \right) \chi_\lambda(\mathbf{r}_2, \mathbf{r}', i\omega), \end{aligned} \tag{3.7}$$

where the exchange-correlation kernel is the functional derivative of the xc potential at coupling strength λ.
The benefit of this construction is not directly obvious; to construct χ_0, we require ψ from the solution of the KS equations and thus already obtaining the correlation energy through an approximation to the xc potential. Additionally, solving Eq. (3.7), requires the xc-kernel which is even more complicated than the xc-potential. The attraction of ACFD is however, that even when setting $f_{xc} = 0$, it yields a nontrivial expression for the correlation energy, namely that of the RPA. Based on the success of the RPA, it is natural to hope that the description of correlation might be further improved by using more sophisticated approximations for f_{xc}.
A natural initial approximation is the local and static adiabatic LDA (ALDA) kernel:

$$f_{xc}^{\mathrm{ALDA}}[n](\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}')f_{xc}^{\mathrm{ALDA}}[n(\mathbf{r})], \tag{3.8}$$
$$\text{with} \quad f_{xc}^{\mathrm{ALDA}}[n(\mathbf{r})] = \frac{d^2}{dn^2}\left(n\epsilon_{xc}^{\mathrm{HEG}}\right)\Big|_{n=n(\mathbf{r})},$$

where $\epsilon_{xc}^{\mathrm{HEG}}$ is the xc energy per electron of the homogeneous electron gas. This approximation is similar to LDA in static DFT but in contrast to LDA, ALDA is not exact for the HEG. The true kernel should depend on both frequencies and position differences. This means that ALDA violates a number of exact conditions, for example it does not decay as $f_{xc} \propto q^{-2}$ for $q \to \infty$. This becomes a problem for total energy calculations as the trace in Eq.(3.6) requires an integral over all $\mathbf{q}$ and an inaccurate description of the response function at large q introduces large errors in the correlation energy. Equivalently, the constant behaviour of f_{xc} at large q makes the pair correlation function divergent at the origin in real space giving rise to convergence problems in ab initio applications. To show these issues in more detail, the coupling constant averaged correlation hole of the HEG is plotted in Fig. 3.1 both in q and real space. The exact representation is from Monte Carlo simulations [62]. Physically, the exchange-correlation hole represents the change in the electron density caused by the presence of an electron. Each electron is surrounded by such an xc hole, resulting in a reduction of its potential energy. The correlation energy per electron is directly related to the coupling constant averaged correlation hole, $\bar{n}_c$:

$$E_c = 2\pi \int_0^\infty \mathrm{d}r r\bar{n}_c(r) = \frac{1}{\pi}\int_0^\infty \mathrm{d}q\bar{n}_c(q),$$

and it is therefore possible to evaluate the effect of f_{xc} on the total energy by analyzing the correlation hole. In RPA, the correlation hole is much too deep for small r which results in an overestimation of the magnitude of the absolute correlation energy. The correlation energy per electron becomes around 0.5 eV too negative for a wide range of densities. By including the exchange part of the ALDA kernel (ALDA$_x$), the q-space representation of the correlation hole is improved but it suffers from a slowly decaying positive tail at large q, resulting in a correlation energy approximately 0.3 eV too large per electron. A better approximation for the total correlation energy is obtained if one cuts off the integral of the correlation hole at $q = 2k_F$ so that $E_c \propto \int_0^{2k_F} \bar{n}_c^{ALDA_x}(q)dq$. This defines the rALDA$_x$ correlation hole, which, as shown in the right part of Fig.3.1 is seen to give a much better description of the short-range part.

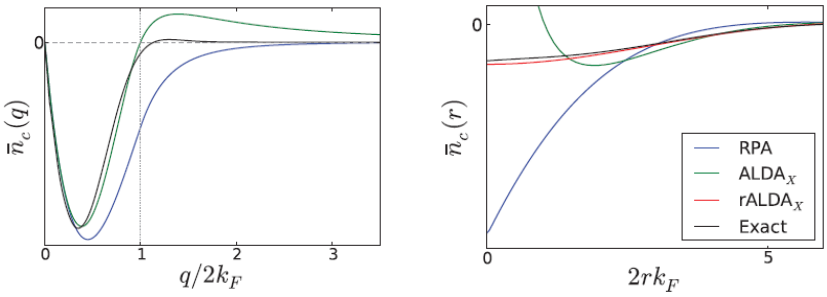


Figure 3.1: Coupling constant averaged correlation hole for the HEG for $r_s = 10$ in q space (left) and real space (right). The rALDA$_x$ is obtained by truncating $\bar{n}_c^{\mathrm{ALDA}_x}(q)$ at its zero point at $2k_F$ (from [54]).

For the HEG, we impose this cutoff in q-space in the Hartree-exchange kernel by

$$f_{Hx}^{\mathrm{rALDA}_x}[n](q) = \theta(2k_F - q) f_{Hx}^{\mathrm{ALDA}_x}[n]$$

which in real space becomes

$$f_{Hx}^{\mathrm{rALDA}_x}[n](r) = \tilde{f}_x^{\mathrm{rALDA}_x}[n](r) + v^r[n](r)$$

with

$$\tilde{f}_x^{\mathrm{rALDA}_x}[n](r) = \frac{f_x^{\mathrm{ALDA}}[n]}{2\pi^2 r^3}\Big[\sin(2k_F r) - 2k_F r\cos(2k_F r)\Big]$$

$$v^r[n](r) = \frac{1}{r}\frac{2}{\pi}\int_0^{2k_F r}\frac{\sin(x)}{x}dx$$

where $k_F = (2\pi^2 n)^{1/3}$. The cutoff in q space is translated into a density dependent width of the delta-function in Eq.(3.8), making it a non-local kernel. The exchange (Hartree) part is shown in the left (right) part of Fig.3.2. The width of the exchange kernel decreases for larger densities (green to blue), approaching the bare exchange kernel and the renormalized Hartree kernel approaches the bare Coulomb kernel for large r. They are both finite at the origin, which is very convenient for numerical evaluation.
One usually applies the xc kernel on top of the exact Hartree kernel so the rALDA exchange kernel is actually defined as:

$$f_x^{\mathrm{rALDA}}[n](r) = \tilde{f}_x^{\mathrm{rALDA}}[n](r) + v^r[n](r) - v(r).$$

This form is desirable for solid state systems where it is usually difficult to converge the long-range tail of $v^r[n](r)$. Indeed,

$$v^r[n](r) - v(r) \to \frac{\sin(2k_F r)}{r} \quad \text{for} \quad r \to \infty,$$

which rapidly averages to zero. Improved convergence is also observed for $\tilde{f}_x^{\mathrm{rALDA}}[n](r)$ due to the $1/r^3$ decay for $r \to \infty$.

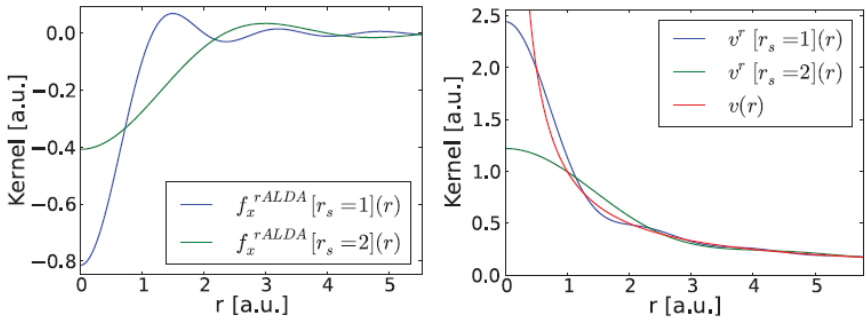
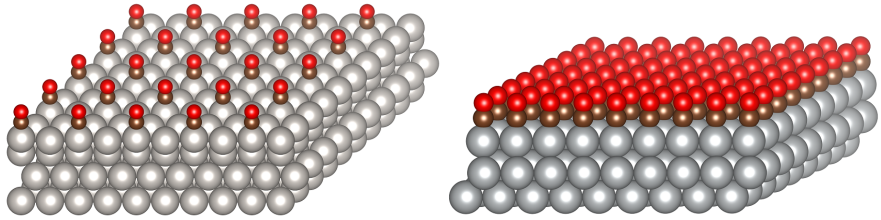


Figure 3.2: Renormalized exchange (left) and Hartree (right) kernel (from [54])

The kernel has been shown to yield more accurate absolute correlation energies and improve atomization and cohesive energies, the dissociation curve of the H_2 molecule, C_6 coefficients and more compared with RPA[53, 54, 55], mainly due to the improved description of short-range correlations. How the same kernel can be used to improve calculations of excitations is discussed in the second part of Chapter 4.

3.3 Publication: Surface chemistry with RPA and rALDA

Benchmark calculations for surface science with the random phase approximation
Per S. Schmidt, Thomas Bligaard and Kristian S. Thygesen
In preparation.



(a) (b)

Figure 3.3: CO adsorbed on Pt(111) at (a) 1/4 and (b) full coverage.

In the attached publication, the RPA and rALDA methods are applied within the topic of surface chemistry. In particular the methods are used to calculate adsorption and surface energies, which are fundamental quantities for predicting and modelling of materials for catalysis.

The application of DFT in surface science is increasing but the expected accuracy across different xc functionals is less established. The direct comparison with experimental adsorption energies is problematic due to the large spread in experimental data for some reactions and the lack thereof for others[23]. It is therefore easier to estimate the accuracy of the various DFT methods by comparing them against higher accuracy computational methods, such as the RPA and rALDA. This, however, requires that we first establish how well these two methods perform for adsorption and surface energies, which brings us back to the problem of comparing with experimental values.

While most previous RPA total energy studies have focused on isolated molecules and bulk solids, there have been some RPA reports on surface adsorption problems[63, 64, 65, 66, 67, 68]. Graphene adsorption on metal surfaces is considered notoriously difficult due to the mixed dispersive-covalent nature of the graphene-metal bond. Nevertheless, the predicted RPA binding distances are in excellent agreement with available experimental data[65].

The RPA has successfully resolved the "CO adsorption puzzle": in contrast to most DFT functionals RPA predicts the correct binding site of CO on Pt(111) and Cu(111). For the case of CO on Pt(111) and Pd(111), Schimka et al.[63] demonstrated how semi-local xc-functionals underestimate the surface energy relative to experiments and at the same time overestimate the adsorption energy. By modifying the xc-functional, either the predicted adsorption energies or surface energies can be improved but always at the expense of the other. In contrast, the RPA improves the description of both properties simultaneously.

CO adsorption on five different surfaces have been investigated in [63] and compared with experiments, showing good agreement. Another renormalized kernel method, the rAPBE, has been applied to the adsorption of CO on Pt(111), with results differing from the RPA by less than 0.05 eV[69]. This further supports that the adsorption of CO is already well described by the RPA. The first part of the attached publication expands the application of the RPA to a wider range of adsorbates and surfaces. RPA is seen to deviate from experiments with 0.2 eV in average, which is similar to the deviation observed with the RPBE xc functional, constructed to improve adsorption energies, and the BEEF-vdW functional which has been fitted to a set of experimental adsorption energies. Even though the mean absolute errors are similar for RPA and the DFT methods, deviations are seen in the individual reactions. The good performance of the RPBE and BEEF-vdW for adsorption energies does not carry over to surface energies, which are hugely underestimated. On the other hand RPA remains accurate also for surface energies.

The experimental adsorption energies are for 1/4 coverage reactions. This makes the unit cell of a corresponding calculation consist of 16-18 atoms, making actual rALDA calculations infeasible. For an illustration of 1/4 coverage, CO adsorbed on Pt is shown in Fig. 3.3(a). For some reactions, the reported experimental values that we compare to differ by up to 0.3 eV, and for others only one reported value was found. This makes the comparison less meaningful.

Two different kinds of adsorption reactions have been investigated. One where the adsorption energy is just a measure of the energy gained when bringing the molecule in contact with the surface of the transition metal slab, for example:

$$\mathrm{CO(g)} + \mathrm{slab} \rightarrow \mathrm{CO/slab}$$

and another where the adsorption energy is a measure of first breaking the bonds of a molecule and subsequently adsorb the individual atoms on the

surface:

$$\frac{1}{2}O_2(g) + \text{slab} \rightarrow O/\text{slab}$$

As mentioned, RPA and rALDA agree for the adsorption of CO on Pt, suggesting that the bonding to the surface is already well described by the RPA.

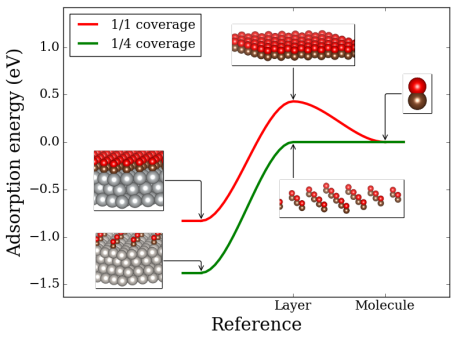
It has however also been shown that the atomization energies calculated with the two methods vary significantly. The atomization energy of O_2 is 4.47 eV with RPA and 5.12 eV with rALDA [54]. The experimental result is 5.20 eV. This deficiency in RPA is related to the poor description of short range correlations[70, 71]. rALDA improves significantly upon RPA in this regard. This will directly influence the adsorption energy of O as that requires dissociating the O_2 molecule. Consequently introducing a rALDA-correction to the RPA adsorption energies from the atomization energy differences is well motivated. For the adsorption of O, the rALDA correction lowers the adsorption energy by half of the atomization energy difference, 0.3 eV, compared to RPA.

In the second part of the publication the rALDA method is used to construct a database of, what we now believe to be high-quality adsorption energies of 200 adsorption reactions. Due to the computational cost of the rALDA method, the database does not consist of adsorption reactions at experimentally relevant coverages. Instead, full coverage reactions were considered with each slab represented by 3 layers in a 1×1 unit cell. Only top-site adsorption was considered on fcc(111) surfaces. CO adsorbed on Pt(111) at full coverage is shown in Fig. 3.3(b). The adsorption site, surface termination or coverage is not experimentally relevant, but that is not a problem as the purpose of the database is to act as a benchmark for various cheaper computational methods.

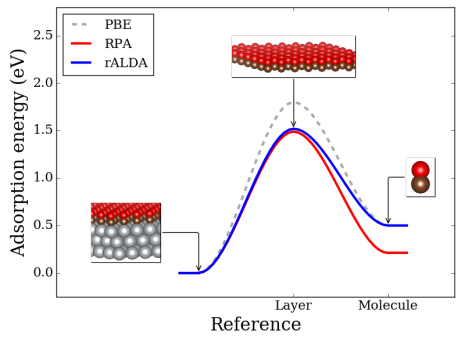
One problem however, is that with full coverage the adsorbate-adsorbate distance is around 2.5 Å and the adsorbate-adsorbate interactions will naturally play a much larger role than was the case for the 1/4 coverage (distances around 5 Å). This suggests that the adsorption reactions consists of two contributions. The first contribution to the adsorption energy is to bring the molecules together to form the full-coverage molecular "layer", and the second contribution is binding this layer to the surface. Each molecule in the layer feels repulsive interactions with the 6 nearest neighbours and the adsorption energy is then naturally lower than for the 1/4 coverage case. One could define the adsorption energy with respect to the molecule layer, but here the reference is kept as the isolated molecule. The adsorption energy of CO on Pt is shown in Fig.3.4(a) at full and 1/4 coverage, calculated with RPA. It

is evident that forming the layer costs around 0.5 eV hence a lower coverage is energetically preferable.

The dissociation of the molecular layer is qualitatively different from the dissociation of a O_2 molecule and the RPA and rALDA methods are found to differ quite significantly for the various adsorbates. If the adsorption energy is defined relative to the molecular layer, RPA and rALDA agree within 0.1 eV across all reactions considered. If the isolated molecule is chosen as the reference, deviations up to 0.42 eV (for NO) is observed between RPA and rALDA, with RPA systematically underbinding. This is shown in Fig.3.4(b) for the adsorption of CO on Mn(111). RPA and rALDA differ by 0.3 eV in the dissociation of the molecular layer. It is evident that PBE accidentally agrees with rALDA even though the adsorption of the molecular layer is overestimated. This is attributed to a fortuitous error cancellation between forming the molecular layer and binding it to the surface.



(a) Adsorption energy of CO on Pt(111) for 1/1 and 1/4 coverage calculated with RPA aligned to zero at the isolated molecule.



(b) Adsorption energy of CO on Mn(111) for 1/1 coverage calculated with RPA and rALDA aligned to zero at the full coverage adsorbate.

Figure 3.4

The difference between RPA and rALDA for the same adsorbate, across different surfaces, was found to be the same, suggesting that the two methods only differ in describing the bonding of the molecular layer.

The database was constructed employing the rALDA method with the gas phase molecule as reference. Across the 200 reactions, the mean absolute error (MAE) of rALDA vs. PBE, RPBE, BEEF-vdW are 0.16, 0.35 and 0.22 eV respectively

One immediate application of the database is to consider the scaling relation between adsorption of O and OH, shown in Fig. 3.5. It has previously been argued that adsorption energies calculated with various GGAs might not be accurate but differences from one system to another should be well produced. This is indeed confirmed from the agreement between the PBE and RPA scaling relations.

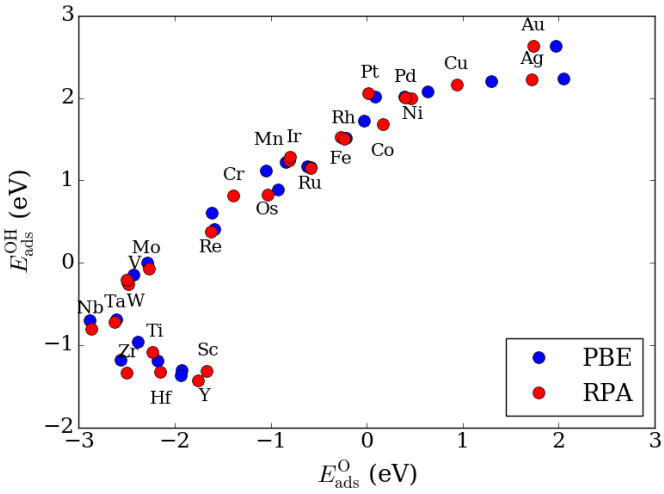


Figure 3.5: Scaling relation between adsorption of O and OH calculated with PBE and RPA.

The database will be useful for guiding construction of new DFT functionals optimized for surface chemistry, capturing the physics present in the rALDA method but keeping the computational speed of DFT. The database will be available at the Computational Materials Repository (`https://cmr.fysik.dtu.dk`).

A computational detail regarding the rALDA method is worth mentioning for future users of the GPAW code. If the adsorption energy is calculated from three systems in the same unit cell; 1) CO/slab, 2) slab and 3) CO molecular layer then both the RPA and rALDA adsorption energies are seen to converge rather quickly with respect to plane wave cutoff and follow the extrapolation scheme of Eq.(3.5) due to error cancellation. If the isolated molecule is the reference, then this must be represented in a unit cell much larger than the molecular layer. The rALDA calculation of the isolated molecule is seen to converge very slowly with plane wave cutoff and not to follow the extrapo-

lation scheme. This is shown in Fig. 3.6 for the N_2 molecule for both RPA and rALDA. The crosses are the extrapolated values obtained using two subsequent cutoff energies. It is evident that the crosses vary significantly for RPA, suggesting that more than two points are necessary to perform the extrapolation. From the rALDA however, it is evident that one cannot apply the extrapolation scheme at all, but one must manually converge the correlation energy. Similar behaviour was seen for all the molecules, requiring plane wave cutoff energies of around 600 eV to be converged with 0.05 eV.

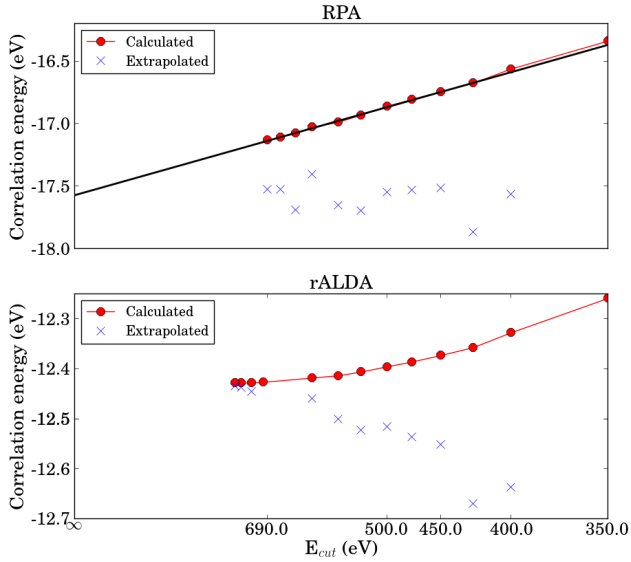


Figure 3.6: Correlation energy vs $1/E_{cut}^{3/2}$ for the N_2 molecule calculated with RPA and rALDA.

3.4 Publication: Electrochemical production of methanol

Electrochemical methanol production from methane
Logi Arnarson, Per S. Schmidt, Kristian S. Thygesen, Ifan E. L. Stephens and Jan Rossmeisl
In preparation.

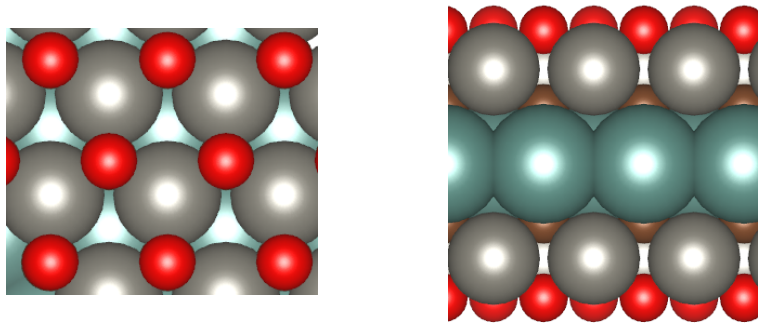


Figure 3.7: $AB_2C_2O_2$ MXene seen from the top and side.

In the attached publication, we investigate electrochemical production of methanol (CH_3OH) from methane (CH_4). Controlling this process can become important in reducing the amount of methane that is burned as waste at oil fields around the globe every day[72]. Methane has 25 times the global warming potential of carbon dioxide, which is why it is usually burned instead of released directly[73]. The burning of methane does however contribute to CO_2 pollution of the atmosphere and an unnecessary waste of the energy stored in the methane. The burning occurs due to a lack in infrastructure allowing for storing and transport of the gaseous methane. It could therefore be attractive to oxidize methane into a liquid fuel such as methanol, which would fit the current infrastructure better.

This process has challenged research in heterogeneous catalysis for decades, leading to inefficient processes requiring high pressures, are unselective or require oxidants such as H_2O_2 which is challenging to transport to remote locations[74, 75]. In contrast, electrochemical oxidation of methane to methanol could be far more attractive:

$$CH_4 + H_2O \rightarrow CH_3OH + H_2, \qquad \Delta G \cong 120\,\mathrm{kj/mol}. \tag{3.9}$$

This reaction can occur in an aqueous environment where an oxygen atom is adsorbed on the surface by splitting water, as is the first part of the oxygen evolution reaction (OER)[76]:

$$\begin{aligned} * + 2H_2O &\rightarrow *OH + H_2O + (H^+ + e^-) && (3.10)\\ &\rightarrow *O + H_2O + 2(H^+ + e^-) && (3.11)\\ &\rightarrow *OOH + 3(H^+ + e^-) && (3.12)\\ &\rightarrow * + O_2 + 4(H^+ + e^-), && (3.13) \end{aligned}$$

The OER can also occur skipping the intermediate step with $*OOH$ and instead have two $*O$ react and form O_2[77]. To form methanol, a methane molecule should come in contact with the adsorbed oxygen:

$$\begin{aligned} *O + CH_4(g) &\rightarrow *O\cdots CH_4 && (3.14)\\ &\rightarrow *OH\cdots CH_3 && (3.15)\\ &\rightarrow *CH_3OH && (3.16)\\ &\rightarrow * + CH_3OH(g) && (3.17) \end{aligned}$$

The main issue as that the methane must react with the oxygen adsorbed on the surface before this oxygen continues the OER and forms O_2 or other products.

In this study we investigate two different surfaces where this process could take place. The (110) surface of rutile transition metal oxides, MeO_2 (Me; a transition metal atom), and monolayer MXenes of the $AB_2C_2O_2$ type (A, B: two different transition metal atoms). MXenes is a new class of two-dimensional materials comprising carbides and nitrides of transition metals. MXenes have shown promise for various applications in lithium/sodium ion batteries and as HER catalysts[78]. MXenes come in various types; M_2X, M_3X_2, M_2XA_2, $M_3X_2A_2$, $M_4X_3A_2$ etc., where M refers to the transition metal, X refers to carbon or nitrogen and A refers to the functional group present on the surface (O, OH or F). In this study, we limit ourselves to the metallic $AB_2C_2O_2$ type, which is shown in Fig.3.7, seen from the top and side. The rationale to use O as the functional group is the solvothermal synthesis of MXenes under acidic condition where O and OH are present to passivate the surface[79]. Previous studies have found that functionalization with O is thermodynamically preferred[80]. The same study investigated the use of some

of the MXenes for HER. They found that the hydrogen adsorption energy varied by up to 0.5 eV depending on the number of metal layers suggesting that catalytic activity can be tuned by controlling layer thickness.
Given the possibility to tune the number of layers and the vast choices of M, X and A, this class of materials is huge. In this study we limit ourselves to seven candidates.

The adsorption energy calculations are carried out with PBE. In the low coverage regime, we know from the previous study that PBE can overbind with up to 0.5 eV. But as we mainly deal with adsorption energy differences, and the scaling relations have been shown to be similar for PBE and the more advanced methods, the use of PBE is justified.

The free energies related to adsorption of O and OH, as relevant for the OERs, are shown in Fig. 3.8 for the seven MXenes together with the ideal catalyst for OER (dashed gray line). For the compounds below the dashed line, we can apply a potential where oxygen binds on the surface but does not spontaneously continue to form O_2 by reaction with another $*O$. It is necessary to calculate the free energy of the $*OOH$ intermediate to determine if it is possible to keep $*O$ on the surface until it can react with CH_4. The O and OH adsorb on the topsite of the transition metal atom.

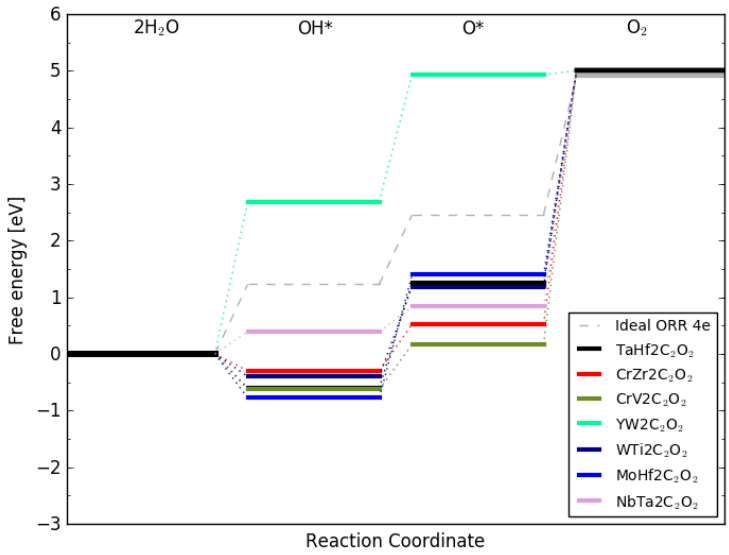


Figure 3.8: OER free energy diagram for seven different MXenes.

Subsequently, it is necessary for the reaction barrier of Eq.(3.14) to (3.17)

to be low enough for the reaction to occur at reasonable temperatures. The reaction pathway of Eq.(3.14) to (3.17) is calculated using the expensive Nudged Elastic Band method with the climbing image variation (CI-NEB). The resulting potential energy landscape along the reaction pathway is shown in Fig. 3.9 for the case of $TaHf_2C_2O_2$ and the activation energy is found to be 0.77 eV. The transition state is $*OH$ on the surface with CH_3 in the vicinity (image 3). This transition state is in consensus with what has been named the radical pathway in a recent publication from Nørskov et al. [81].

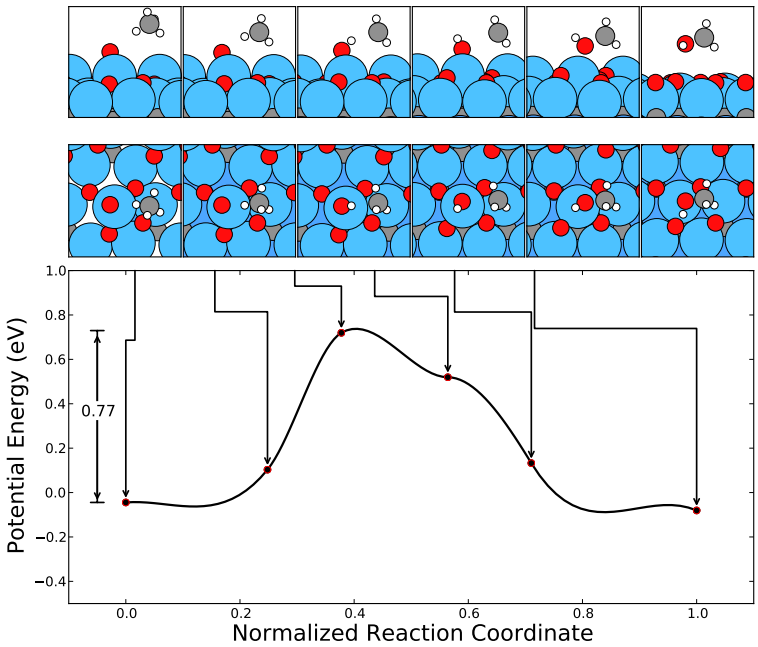


Figure 3.9: Reaction pathway producing methanol from methane on $TaHf_2C_2O_2$ from a CI-NEB calculation.

This pathway suggests that the energy required to break the first C-H bond, to activate methane, equals the energy needed to deliver a single H atom to the surface and form CH_3 from CH_4, i.e.:

$$E_{\text{act}}^{\text{estimated}} = \left(E_{\text{CH}_3(\text{g})} - E_{\text{CH}_4(\text{g})}\right) + \left(E_{*\text{OH}} - E_{*\text{O}}\right) \tag{3.18}$$

This corresponds to converting the sp3 hybridisation in CH_4 to sp2 hybridisation in CH_3 and the double bonded oxygen to the metal atom to a single bonded OH group on the surface. In Fig. 3.10 we therefore report the activation energy calculated from Eq. (3.18) versus $E_{*\mathrm{OH}} - E_{*\mathrm{O}}$. Due to the computational cost, CI-NEB calculations were only carried out for a few materials, here shown with squares. In general the resulting CI-NEB calculations agree well with the simple model.

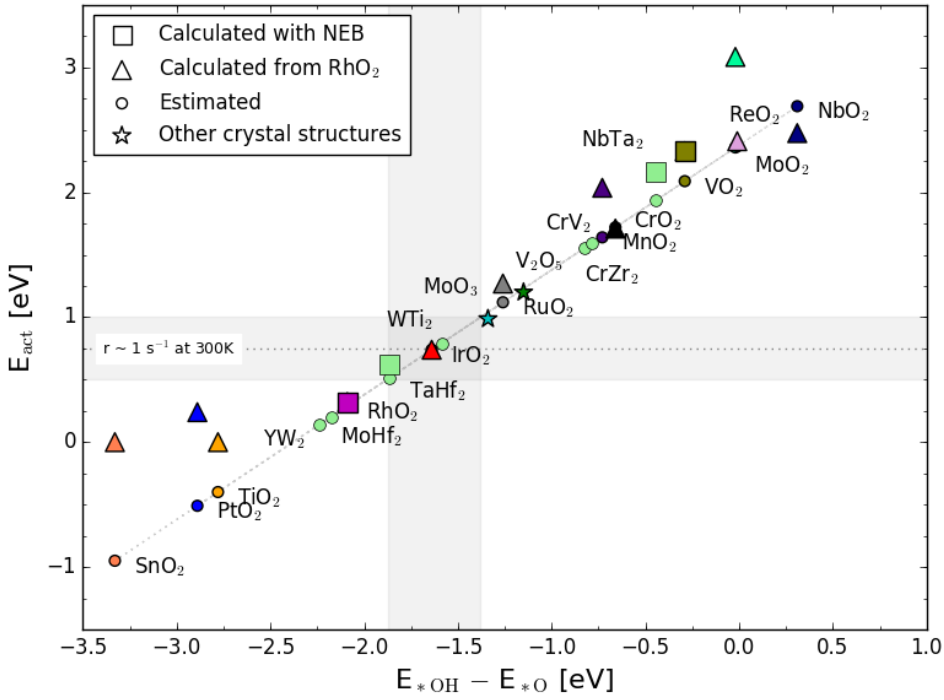


Figure 3.10: Activation energy shown versus the descriptor for both rutile (110) MeO_2 structures and MXenes (in light green). The activation energies calculated with CI-NEB (squares) agree well with the simple model.

As a rule of thumb for gas phase chemistry to obtain a rate of $\sim$1 s^{-1} at 300K the energy barrier should be $\sim$0.75 eV. A few materials fall at or below this value. To ensure high selectivity towards CH_3OH instead of forming O_2, only a moderate potential should be applied, according to the full OER pathway. This will come at the cost of a higher activation energy. With this in mind, promising candidates could be $MoHf_2C_2O_2$ and $TaHf_2C_2O_2$ since the

reaction barriers are close to or below 0.75 eV and the potential necessary to adsorb oxygen on the surface is not sufficient for the oxygen to continue to form O_2 by reacting with another $*O$. The free energy of the $*OOH$ intermediate remains to be calculated.

CHAPTER 4

Excitations: The GW Approximation

DFT is constructed to describe ground state properties. Therefore it is not surprising that the KS eigenvalues provide a poor description of excitations. In the derivation of Kohn-Sham theory, the eigenvalues $\epsilon_i^{\mathrm{DFT}}$ enter as Lagrange multipliers ensuring orthogonality between the eigenfunctions ϕ_i^{DFT} of the fictitious non-interacting particles. $\epsilon_i^{\mathrm{DFT}}$ and ϕ_i^{DFT}, as such, are simply mathematical tools and the only relevant physical information they invoke is that the square of the eigenfunctions sum up to the exact ground state density and that the highest occupied eigenvalue equals the negative of the ionization potential[30].

A systematic way to access the excited states, hence overcoming some of the limitations of DFT, is provided by many-body perturbation theory (MBPT). In MBPT, instead of dealing with non-interacting particles as in the KS problem, a quasiparticle formalism is employed where the electrons are dressed with the interaction of the surroundings. In particular, the repulsive Coulomb interaction between electrons leads to a depletion of negative charge around a given electron. The electron and its surrounding positive screening charge forms the so-called quasiparticle. The single-particle Green's function is the foundation for the mathematical description of quasiparticles and in order to determine the Green's function, complete knowledge of the quasiparticle self-energy is required. This is a non-Hermitian, non-local, energy dependent operator which incorporates exchange and correlation effects beyond the Hartree approximation. The self-energy must therefore be approximated[82, 83]. In this chapter, various methods are discussed and applied to the calculation of excitations in bulk and 2D semiconductors and insulators.

4.1 The quasiparticle equation

The success of the methods presented in the following chapter is based on the fact that weakly interacting quasiparticles can be used to describe excitations of a system of strongly interacting electrons. The bare electron and its positive screening charge form the quasiparticle which interacts weakly with other quasiparticles through the screened, rather than the bare, Coulomb potential. The interaction leads to complex eigenenergies whose imaginary part is inversely proportional to the quasiparticle life time. This is finite since quasiparticles are only approximate eigenstates of the electronic Hamiltonian[82]. The energy difference between the bare particle and the quasiparticle is described by the self-energy, Σ, accounting for all exchange and correlation effects, and the central equation governing the behaviour of the quasiparticles, is the quasiparticle-equation:

$$\left[-\frac{1}{2}\nabla^2 + v_{\mathrm{Ha}}(\mathbf{r}) + v_{\mathrm{ext}}(\mathbf{r})\right]\phi_i^{\mathrm{QP}}(\mathbf{r}) + \int d\mathbf{r}'\Sigma_{\mathrm{xc}}(\mathbf{r},\mathbf{r}',\epsilon_i^{\mathrm{QP}})\phi_i^{\mathrm{QP}}(\mathbf{r}') = \epsilon_i^{\mathrm{QP}}\phi_i^{\mathrm{QP}}(\mathbf{r}) \tag{4.1}$$

where $v_{\mathrm{Ha}}(\mathbf{r})$ is the Hartree potential, $v_{\mathrm{ext}}(\mathbf{r})$ the external potential and ϵ_i^{QP}, $\phi_i^{\mathrm{QP}}(\mathbf{r})$ the quasiparticle energies and wave functions respectively. The equation has the structure of a single-particle Schrödinger equation and with a local and static approximation to the self-energy, $\Sigma_{\mathrm{xc}}(\mathbf{r},\mathbf{r}',\epsilon_i^{\mathrm{QP}}) = v_{xc}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}')$, one obtains the Kohn-Sham equation.

Experimentally, quasiparticle energies and lifetimes are measured by direct or inverse photoemission where an electron is added or removed from the system. In direct photoemission, a photon with energy $h\nu$ enters the sample and an electron in the valence band absorbs the photon energy. The electron leaves the sample with a given kinetic energy, E_{kin}, which can be measured directly. The energy of the valence band state relative to vacuum is found from the simple relation $E_{\mathrm{kin}} = E_{\mathrm{VB}} + h\nu$. In inverse photoemission, an electron with kinetic energy E_{kin} is injected into the solid, where it will lose energy through photon emission before coming to a rest in the conduction band. The energy of the emitted photon, $h\nu$ is measured and the energy of the conduction band state determined from $E_{\mathrm{kin}} = E_{\mathrm{CB}} + h\nu$. This means that we can readily compare calculated QP energies for the conduction (valence) band with E_{CB} (E_{VB})[83].

Solving the QP equation is complicated by the fact that the self-energy must be evaluated at the QP energies which are not known a priori. Instead, assuming that the KS equation is a good description of the system, one can cal-

culate the correction to the KS eigenvalues by first order perturbation theory. Adding and subtracting $v_{xc}^{\mathrm{DFT}}(\mathbf{x})$ to the QP equation as follows

$$\left[-\frac{1}{2}\nabla^2 + v_{\mathrm{Ha}}(\mathbf{r}) + v_{\mathrm{ext}}(\mathbf{r}) + v_{xc}^{\mathrm{DFT}}(\mathbf{r})\right]\phi_i^{\mathrm{QP}}(\mathbf{r}) + \int d\mathbf{r}'\left[\Sigma^{\mathrm{xc}}(\mathbf{r},\mathbf{r}',\epsilon_i^{\mathrm{QP}}) - v_{xc}^{\mathrm{DFT}}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}')\right]\phi_i^{\mathrm{QP}}(\mathbf{r}') = \epsilon_i^{\mathrm{QP}}\phi_i^{\mathrm{QP}}(\mathbf{r})$$

and assuming $\Sigma_{\mathrm{xc}} - v_{\mathrm{xc}}$ to be small, one can obtain the self-energy from a first order expansion around the QP eigenvalue

$$\Sigma_{\mathrm{xc}}(\epsilon_i^{\mathrm{QP}}) \approx \Sigma_{\mathrm{xc}}(\epsilon_i^{\mathrm{DFT}}) + (\epsilon_i^{\mathrm{QP}} - \epsilon_i^{\mathrm{DFT}})\frac{\partial\Sigma_{\mathrm{xc}}(\omega)}{\partial\omega}\bigg|_{\omega=\epsilon_i^{\mathrm{DFT}}}$$

The expectation value of the QP equation then becomes

$$\epsilon_i^{\mathrm{QP}} = \epsilon_i^{\mathrm{DFT}} + Z_i\,\langle\phi_i^{\mathrm{DFT}}|\Sigma_{\mathrm{xc}}(\epsilon_i^{\mathrm{DFT}}) - v_{\mathrm{xc}}|\phi_i^{\mathrm{DFT}}\rangle \tag{4.2}$$

where

$$Z_i = \left\langle \phi_i^{\mathrm{DFT}}\middle|1 - \frac{\partial\Sigma^{\mathrm{xc}}(\omega)}{\partial\omega}\bigg|_{\omega=\epsilon_i^{\mathrm{DFT}}}\middle|\phi_i^{\mathrm{DFT}}\right\rangle^{-1}$$

The renormalization factor, Z, approximates the true QP norm and if $Z \ll 1$ we can conclude that ϕ_i^{DFT} is not a proper QP state. Either the electrons are strongly correlated and the QP picture does not apply or the DFT orbital is not a good approximation to the true QP wave function[84]. It has been observed that $Z \approx 0.7 - 0.9$ for most semiconductors considered in this work.

To allow for such calculations, a suitable approximation for the self-energy is necessary. Such an approximation will be derived in the following[33, 82, 83]. The one-particle Green's function, also known as the single-particle propagator, is defined as:

$$G(\mathbf{r}t,\mathbf{r}'t') = -i\,\langle N|T[\hat{\psi}(\mathbf{r}t)\hat{\psi}^\dagger(\mathbf{r}'t')]|N\rangle$$

where T is the time-ordering operator, $\hat{\psi}(\mathbf{r}t)$ ($\hat{\psi}^\dagger(\mathbf{r}t)$) the annihilation (creation) operator and $|N\rangle$ the N-electron ground state. For $t > t'$, G describes the propagation of a particle added to the system at time t' propagating in the material until time t. With $E_{N,i}$, the energy of the N-electron system in

state i, we define the quasiparticle wave functions and energies as

$$\epsilon_i^{\mathrm{QP}} = \begin{cases} E_{N+1,i} - E_{N,0} & \epsilon_i^{\mathrm{QP}} \geq \mu \\ E_{N,0} - E_{N-1,i} & \epsilon_i^{\mathrm{QP}} \leq \mu \end{cases}$$
$$\phi_i^{\mathrm{QP}}(\mathbf{r}) = \begin{cases} \langle N|\hat{\psi}(\mathbf{r})|N+1,i\rangle & \epsilon_i^{\mathrm{QP}} \geq \mu \\ \langle N-1,i|\hat{\psi}(\mathbf{r})|N\rangle & \epsilon_i^{\mathrm{QP}} \leq \mu \end{cases}$$

where μ is the chemical potential. By inserting a set of complete states and Fourier transforming to frequency space, one can identify the quasiparticle energies as poles of the Green's function:

$$G(\mathbf{r},\mathbf{r}',E) = \sum_i \frac{\phi_i^{\mathrm{QP}}(\mathbf{r})[\phi_i^{\mathrm{QP}}(\mathbf{r}')]^*}{E - \epsilon_i^{\mathrm{QP}}}$$

The interacting Green's function can also be determined by introducing the quasiparticle self-energy via its Dyson equation [85]

$$G(\mathbf{r},\mathbf{r}',E) = G_0(\mathbf{r},\mathbf{r}',E) + \int\int d\mathbf{r}_1 d\mathbf{r}_2 G_0(\mathbf{r},\mathbf{r}_1,E)\Sigma_{xc}(\mathbf{r}_1,\mathbf{r}_2,E)G(\mathbf{r}_2,\mathbf{r}',E)$$

where $G_0(\mathbf{r},\mathbf{r}',E)$ is the Green's function of the non-interacting system (the Hartree Green's function, $\Sigma_{xc} = 0$). From here on the xc subscript of the self-energy is omitted.
It can be shown that the self-energy can be obtained from a closed set of equations, known as Hedin's equations[86, 87, 82]. Together with the Dyson equation above, they link the single-particle propagator G, the self-energy Σ, the screened interaction W to the irreducible polarizability P and the vertex function Γ. P (Γ) describes the change in the density n (inverse Green's function G^{-1}) upon a change in the total potential:

$$P(1,2) = \frac{\delta n(1)}{\delta v(2)}$$
$$\Gamma(1,2,3) = -\frac{\delta G^{-1}(1,2)}{\delta v(3)}$$

where 1,2,... are shorthand notation for position, spin and time variables

$(1 = \mathbf{r}_1, \sigma_1, t_1)$. With the given definitions, Hedin's equations read

$$\Sigma(1,2) = i \int d(3,4) W(1^+,3) G(1,4) \Gamma(4,2,3) \tag{4.3}$$

$$P(1,2) = -i \int d(3,4) G(2,3) G(4,2) \Gamma(3,4,1) \tag{4.4}$$

$$W(1,2) = v(1,2) + \int d(3,4) v(1,3) P(3,4) W(4,2) \tag{4.5}$$

$$\Gamma(1,2,3) = \delta(1,2)\delta(1,3) + \int d(4,5,6,7) \frac{\delta\Sigma(1,2)}{\delta G(4,5)} G(4,6) G(7,5) \Gamma(6,7,3) \tag{4.6}$$

$$\tag{4.7}$$

1^+ implies that the time variable is augmented by a positive infinitesimal. Because of the interdependence of the five equations, the self-energy must be found self-consistently, i.e. starting with an approximation for one or more quantities and then iterate Hedin's equations until self-consistency, as sketched in Fig. 4.1. In practice, due to computational limitations, the process is usually terminated after one iteration.

In the standard and widely used GW approximation, the iteration of Hedin's equations start from $\Sigma^{(0)} = 0$ in Eq.(4.6) resulting in the following expressions (omitting position-spin-time variables):

$$\begin{aligned} \Sigma^{(0)} &= 0 \\ \Gamma &= 1 \\ P &= -iGG = \chi_0 \\ W &= v + v\chi_0 W \Rightarrow \\ W &= v(1 - v\chi_0)^{-1} \\ G &= G_0 \\ \Sigma^{(1)} &= iG_0 W \end{aligned}$$

or in real-space and frequency domain

$$\Sigma(\mathbf{r}, \mathbf{r}', \omega) = \frac{i}{2\pi} \int d\omega' G_0(\mathbf{r}, \mathbf{r}', \omega + \omega') W(\mathbf{r}, \mathbf{r}', \omega') e^{i\omega'\eta} \tag{4.8}$$

$$\text{and} \quad W(\mathbf{r}, \mathbf{r}', \omega) = \int d\mathbf{r}'' \epsilon^{-1}(\mathbf{r}, \mathbf{r}'', \omega) v(\mathbf{r}'', \mathbf{r}')$$

$$\text{with} \quad \epsilon^{-1}(\mathbf{r}, \mathbf{r}', \omega) = \delta(\mathbf{r} - \mathbf{r}') + \int d\mathbf{r}'' v(\mathbf{r}, \mathbf{r}'') \chi_0(\mathbf{r}'', \mathbf{r}', \omega)$$

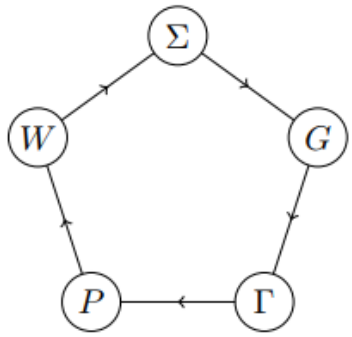


Figure 4.1: GW iteration scheme.

Starting from the Hartree approximation, the resulting G and W should be evaluated using Hartree eigenenergies and eigenfunctions as input. This, however, turns out to be a poor starting point, since the Hartree approximation fails dramatically in describing semiconductors and insulators. LDA or GGA eigenenergies and eigenfunctions have been shown to be a better approximation for the QP energies and wave functions and therefore represent a better starting point when adopting the "best G, best W" philosophy[88]. This philosophy is fundamentally unsatisfactory but evaluating the self-energy in a one-shot manner on top of DFT wave functions has shown to be very successful at describing electronic excitations for a large variety of materials, although systematically underestimating band gaps[89, 90]. One attempt at addressing the starting point dependence is self-consistency as briefly discussed in Section 4.1.2. Another way of justifying a LDA starting point is presented in Section 4.3. In the same section, the effect of including a non-trivial vertex, Γ, on the resulting QP energies is addressed.

4.1.1 Implementation of the GW method in GPAW

In GPAW the self-energy is calculated in a plane wave basis

$$\begin{aligned}\Sigma_{n\mathbf{k}}(\epsilon_{n\mathbf{k}}^{\mathrm{DFT}}) &= \langle n\mathbf{k}|\Sigma(\omega=\epsilon_{n\mathbf{k}}^{\mathrm{DFT}})|n\mathbf{k}\rangle \\ &= \frac{1}{\Omega}\sum_{\mathbf{G}\mathbf{G}'}\sum_{\mathbf{q}}^{1.\mathrm{BZ}}\sum_{m}^{\mathrm{all}}\frac{i}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}\omega' W_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega') \\ &\times \frac{\rho_{m\mathbf{k}-\mathbf{q}}^{n\mathbf{k}}(\mathbf{G})\left[\rho_{m\mathbf{k}-\mathbf{q}}^{n\mathbf{k}}(\mathbf{G}')\right]^*}{\epsilon_{n\mathbf{k}}^{\mathrm{DFT}}-\epsilon_{m\mathbf{k}-\mathbf{q}}^{\mathrm{DFT}}-\omega'+i\eta\mathrm{sgn}(\epsilon_{m\mathbf{k}-\mathbf{q}}-\mu)}\end{aligned} \tag{4.9}$$

where $\rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G}) = \langle n\mathbf{k}|e^{i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}|m\mathbf{k}-\mathbf{q}\rangle$, termed a pair-density. The self-energy is split into an exchange and a correlation term,

$$\begin{aligned}\Sigma &= \Sigma_x + \Sigma_c \\ &= iGv + iG(\epsilon^{-1}v - v) \\ &= iGv + iG\overline{W}\end{aligned}$$

The correlation part of the self-energy, Σ_c, is calculated from the screened Coulomb interaction

$$\begin{aligned}\overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) &= \sqrt{v_{\mathbf{G}}(\mathbf{q})}\left[\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) - \delta_{\mathbf{G}\mathbf{G}'}\right]\sqrt{v_{\mathbf{G}'}(\mathbf{q})} \\ &= \frac{4\pi}{|\mathbf{q}+\mathbf{G}|}\left[\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) - \delta_{\mathbf{G}\mathbf{G}'}\right]\frac{1}{|\mathbf{q}+\mathbf{G}'|},\end{aligned}$$

where the *symmetrized* bare Coulomb interaction is used. As evident from Eq.(4.9), the calculation of the matrix-elements of the self-energy requires the summation over the entire first Brillouin zone, $\sum_{\mathbf{q}}^{1.\mathrm{BZ}}$, including the Γ point. This gives rise to some numerical problems as the expression for $\overline{W}$ is seen to diverge for $\mathbf{G},\mathbf{q} = 0$ and $\mathbf{G}',\mathbf{q} = 0$, also known as the head and wings of $\overline{W}$. The divergence is however integrable and in practice the $\mathbf{q}$-sum is split into two parts

$$\sum_{\mathbf{q}}^{1.\mathrm{BZ}} S(\mathbf{q}) \to \sum_{\mathbf{q}\neq\mathbf{0}}^{1.\mathrm{BZ}} S(\mathbf{q}) + \frac{1}{\Omega_0}\int \mathrm{d}^3\mathbf{q} S(\mathbf{q})$$

where $\Omega_0 = \frac{1}{N_q}\frac{(2\pi)^3}{V_{\mathrm{cell}}}$ is the volume of the first Brillouin zone attributed to the Γ point, namely $1/N_q$ of the entire volume, Ω_0. Everything but the screened potential is assumed constant inside Ω_0 so the integral left to evaluate is simply (for the head)

$$\frac{1}{\Omega_0}\int \mathrm{d}^3\mathbf{q} S(\mathbf{q}) = G_{00}(0,\omega)\frac{1}{\Omega_0}\int_{\Omega_0} \mathrm{d}^3\mathbf{q}\overline{W}_{00}(\mathbf{q},\omega)$$

In 3D, the dielectric function goes to a finite constant for $q \to 0$ which allows us to simplify the integral to

$$\frac{1}{\Omega_0}\int_{\Omega_0} \mathrm{d}^3\mathbf{q}\, W_{00}(\mathbf{q},\omega) = \frac{1}{\Omega_0}\left[\epsilon^{-1}_{\mathbf{00}}(0,\omega) - 1\right]\int_{\Omega_0} \mathrm{d}^3\mathbf{q}\,\frac{4\pi}{|\mathbf{q}|^2} \tag{4.10}$$

The integral on the RHS can be solved analytically for a spherical volume of radius q_r defined to have the same volume as Ω_0 ($4/3\pi q_r^3 = \Omega_0$):

$$\begin{aligned}\frac{1}{\Omega_0}\int_{\Omega_0} \mathrm{d}^3\mathbf{q}\overline{W}_{00}(\mathbf{q},\omega) &= \left[\epsilon^{-1}_{\mathbf{00}}(0,\omega)-1\right]\frac{1}{\Omega_0}(4\pi)^2 q_r \\ &= \left[\epsilon^{-1}_{\mathbf{00}}(0,\omega)-1\right]16\pi^2\left(\frac{3}{4\pi}\right)^{1/3}\Omega_0^{-2/3}\end{aligned}$$

A similar expression is found for the wings.

One evident problem relevant to this thesis is that the same treatment cannot be applied to 2D systems. For 2D systems the screened potential has a $\mathbf{q}$-dependent behaviour that is qualitatively different than for 3D systems. First off, it is not justified setting the dielectric function to a constant inside Ω_0 and secondly, there is a need of replacing the standard Coulomb potential with a truncated one to avoid artificial interactions between neighbouring layers in our supercell calculations. Therefore it has been necessary to implement a new way of treating the $\mathbf{q} = 0$ contribution to the self-energy, as will be explained in greater detail in Sec. 4.2 and the attached publication.
The divergence of the Coulomb interaction is present in the exchange part of the self-energy as well, where it can be integrated in a similar fashion. The best performance (fastest convergence with respect to number of $\mathbf{q}$-points included) is found to be achieved when using a Wigner-Seitz truncated Coulomb interaction[91], which is the method employed in GPAW.

To future users of the GW implementation in the GPAW code, it is worth knowing that possible improvements and features, not presently implemented in GPAW, has been presented in the literature over the course of this project. If added to the GPAW code, these could improve upon computational speed, memory requirements and expand the systems for which it is feasible to carry out GW calculations. The most significant additions have been listed below.

Possible improvements to the GW implementation:

1. Frequency integration: Currently carried out on the real axis using a logarithmic grid. This could be transformed to the imaginary axis using contour integration, reducing the amount of points required saving both time and memory [92].
2. The number of bands included in the summation over unoccupied bands in the construction of χ_0 and again of $\langle n\mathbf{k}|\Sigma^{\mathrm{xc}}|n\mathbf{k}\rangle$ can be reduced significantly by exploiting the $\sqrt{E}$-behaviour of the DOS[93].
3. The summation over unoccupied bands can be completely omitted by use of the Sternheimer equation, a Lanczos algorithm and contour deformation techniques, see [94].
4. Expand the code to be applicable to metallic and spin-polarized systems.

4.1.2 Eigenvalue self-consistency

The one-shot approach of evaluating the GW self-energy using DFT input orbitals has yielded results in good agreement with experiments [89, 90]. Part of the success has been attributed to a fortuitous cancellation of errors introduced by the shortcomings of the G_0W_0 method itself and those introduced by the approximation to the exchange and correlation (xc) potential used in the preceding DFT calculation. The agreement is however not perfect and G_0W_0 is known to systematically underestimate band gaps. One attempt at improving the agreement with experiments is to iterate the QP equation, Eq.4.1, until self-consistency, starting from DFT input orbitals and eigenvalues and ending up with the QP equivalent. A simple step between G_0W_0 and full self-consistent GW is obtained by only iterating the eigenvalues, keeping the wave functions fixed at the DFT level. LDA wave functions are believed to be reasonable for many semi-conductors but fail in describing transition-metal oxides or rare-earth oxides. The eigenvalues can be updated in G only (termed GW_0) or in both G and W. Both methods lead to an opening of the band gap by a symmetric upward (downward) shift of the conduction (valence) band. GW_0 is seen to bring band gaps in closer agreement with experiments where as eigenvalue-GW is seen to overestimate them significantly[95]. Eigenvalue

self-consistency has been implemented in GPAW during this project.

The eigenvalues of the $N+1$ iteration toward self-consistency is in principle calculated from

$$\epsilon^{N+1} = \mathrm{Re}\Big[\langle\psi|T + v_H + \Sigma(\epsilon^{N+1})|\psi\rangle \Big]$$

but by linearising $\Sigma(\epsilon^{N+1})$ around $\Sigma(\epsilon^N)$ we obtain

$$\epsilon^{N+1} = \mathrm{Re}\Big[\langle\psi|T + v_H + \Sigma(\epsilon^N)|\psi\rangle \Big] + (\epsilon^{N+1} - \epsilon^N)\mathrm{Re}\Big[\langle\psi|\frac{\partial\Sigma(\omega)}{\partial\omega}\Big|_{\omega=\epsilon^N}|\psi\rangle \Big] \Rightarrow$$

$$\epsilon^{N+1} = \epsilon^N + Z^N\mathrm{Re}\Big[\langle\psi|T + v_H + \Sigma(\epsilon^N)|\psi\rangle - \epsilon^N \Big]$$

In the current implementation of GW_0 it is possible to store W_0 after the first iteration, making subsequent iterations computationally faster to evaluate, but requiring a significant amount of disk space. It has been observed that a smoother convergence of the eigenvalues is achieved when the eigenvalues of the $N+1$ iteration is mixed with the previous eigenvalues before continuing. By default the mixing is 50%.

From Eq.(4.9) it is evident that the self-energy requires a summation over all bands, $\approx 1000 - 3000$ for typical systems. When iterating the GW_0 method, it is thus in principle required that the shift in all bands are calculated. One detail of the current implementation worth mentioning is that it is currently possible to approximate the shifts of all the unoccupied bands by the calculated shift of the conduction band, averaged in $\mathbf{k}$. Likewise the downward shift of all the occupied bands can be approximated by the $\mathbf{k}$-averaged shift of the valence band.

4.2 Publication: Optimizing GW for 2D materials

Efficient many-body calculations of two-dimensional materials using exact limits for the screened potential: Band gaps of MoS_2, h-BN and phosphorene

Filip A. Rasmussen, Per S. Schmidt, Kirsten T. Winther and Kristian S. Thygesen

Phys. Rev. B 94, 155406 - Published 6 October 2016

When carrying out GW calculations of 2D materials, the evaluation of the dielectric function and the screened Coulomb interaction require a different numerical strategy compared to the 3D case. The complications introduced by the low dimensionality in the evaluation of the dielectric function are most easily presented considering a simple strict 2D model of a homogeneous and isotropic semiconductor. For $\mathbf{G} = 0$, in the small q limit, the pair densities, $\rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G})$ become

$$\rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G}) = \langle \psi_{n\mathbf{k}} | e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} | \psi_{m\mathbf{k}-\mathbf{q}} \rangle \to -i\mathbf{q}\cdot\frac{\langle \psi_{n\mathbf{k}} | \nabla | \psi_{m\mathbf{k}} \rangle}{\epsilon_{m\mathbf{k}} - \epsilon_{n\mathbf{k}}}, \quad \text{for} \quad \mathbf{G} = 0, \mathbf{q} \to 0$$

and therefore the density response function, Eq. (3.3), takes the form $\chi_0^{2\mathrm{D}}(q) = -\alpha_{2\mathrm{D}} q^2$, where $\alpha_{2\mathrm{D}}$ is the 2D polarizability[18]. The 2D Fourier transform of $1/r$ is $2\pi/q$ and thus the leading order of the dielectric function becomes

$$\epsilon^{2\mathrm{D}}(q) = 1 - v(q)\chi_0^{2\mathrm{D}}(q) = 1 + 2\pi\alpha_{2\mathrm{D}} q$$

and the correlation part of the screened Coulomb interaction is then

$$\overline{W}^{2\mathrm{D}}(q) = v(\epsilon^{-1} - 1) = \frac{2\pi}{q}\left(\frac{1}{1 + 2\pi\alpha_{2\mathrm{D}} q} - 1\right) = -\frac{4\pi^2\alpha_{2\mathrm{D}}}{1 + 2\pi\alpha_{2\mathrm{D}} q}$$

which is seen to take a finite value for $q = 0$, as opposed to the 3D case where the correlation part of the screened interaction diverges. Now if one was to calculate $\overline{W}^{2\mathrm{D}}(q)$ with the standard 3D approach, one would obtain $\epsilon^{2\mathrm{D}}(q \to 0) = 1$ and consequently $\overline{W}^{2\mathrm{D}}(q)$ in Eq.(4.10) would be zero. From this, it is clear that the entire q-dependence around $q = 0$ of the 2D dielectric function in Eq.(4.10) should be taken into account when evaluating $\int_{\Omega_0} d^3\mathbf{q}\, \overline{W}^{2\mathrm{D}}(\mathbf{q}, \omega)$.

The static dielectric function of a range of 2D materials is shown in Fig. 4.2(a) as a function of q. The linear behaviour around $q = 0$ is clearly observed, but the slope varies a lot between the different materials.

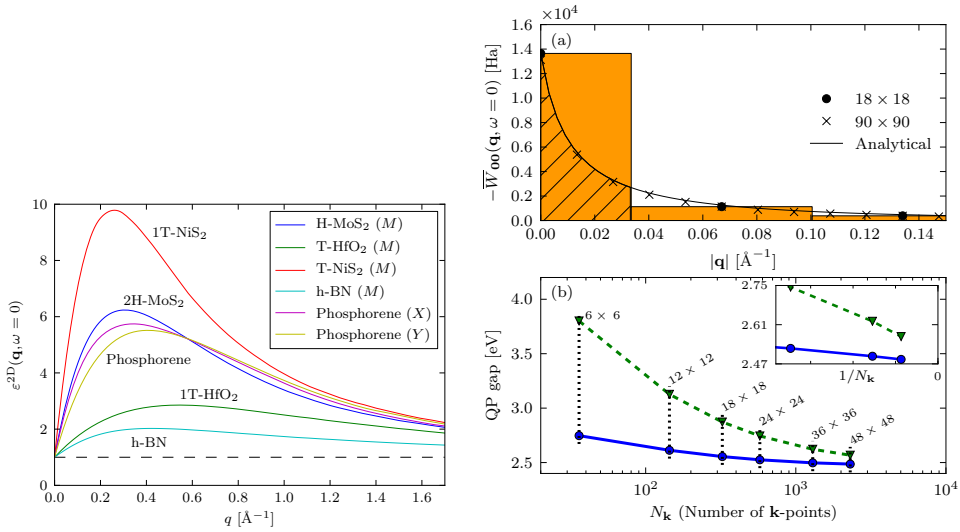


Figure 4.2: (a) Dielectric functions for a range of 2D materials as a function of q. (b) Quasiparticle band gap of 2H-MoS_2 as a function of number of $\mathbf{k}$-points included.

Computationally, 2D materials are represented in an elongated unitcell with 10-15 Å of vacuum in the out-of-plane direction isolating the layer from the artificial neighbours generated by the use of periodic boundary conditions in all three directions. Using the standard Coulomb interaction, as done for calculations of bulk materials, leads to significant over-screening due to the long range interaction between the repeated images. This can be solved by using very large unit cells but this increases the computational time dramatically. To avoid having to use very large unit cells in the out-of-plane direction and at the same time avoid interactions between neighbouring layers, the Coulomb potential is truncated in the out-of-plane direction.

$$v^{\text{trunc}}(\mathbf{r}, z; \mathbf{r}', z') = \frac{\theta(R_c - |z' - z|)}{\sqrt{(\mathbf{r}' - \mathbf{r})^2 + (z' - z)^2}}$$

The truncation distance is chosen to be half the unitcell size in the out-of-plane direction, $R_c = L/2$, to arrive at the following expression in reciprocal space [96, 97]

$$v^{\text{trunc}}_{\mathbf{G}}(\mathbf{q}_\parallel) = \frac{4\pi}{|\mathbf{q}_\parallel + \mathbf{G}|^2}\left[1 - \exp^{-|\mathbf{q}_\parallel + \mathbf{G}_\parallel| L/2} \cos(G_z L/2)\right].$$

With the current implementation of this method, the layer should be centered in the unit cell. We recover the usual untruncated 3D Coulomb potential, $v(\mathbf{q}) = \frac{4\pi}{\mathbf{q}^2}$ for $|\mathbf{q}_\parallel + \mathbf{G}_\parallel|L/2 \gg 1$, i.e. in the small wavelength limit. In the long wavelength limit, $\mathbf{G} = 0$, $q_z = 0$, $\mathbf{q}_\parallel \to 0$, the truncated interaction becomes $v_0^{\mathrm{2D}}(\mathbf{q}_\parallel) \approx \frac{2\pi L}{\mathbf{q}_\parallel}$ which is reduced by a power of $\mathbf{q}$ compared to that of the full Coulomb interaction. This has important consequences for the form of the screened interaction in calculations of 2D materials.

In the attached publication, the truncated Coulomb interaction and the 2D dielectric function is combined to derive an expression for the head, wings and body of $\overline{W}$ in the small $\mathbf{q}$ limit. The method is applied to three different 2D semiconductors to show how it significantly improves the k-points convergence, enabling large-scale GW calculations without compromising accuracy. The improvements is shown in Fig. 4.2(b): The upper figure shows how poorly the rough discrete sampling (18×18 k-pts) represents the true $\overline{W}$. On the bottom figure it is shown how the quasiparticle gap converges with respect to the number of $\mathbf{q}$-points, both with and without the analytic correction to the $q = 0$ term. The band gap is converged to within 0.1 eV using $12 \times 12 \times 1$ $\mathbf{q}$-points when applying the correction whereas at least $36 \times 36 \times 1$ $\mathbf{q}$-points were necessary without.

4.3 Publication: Including vertex corrections in GW calculations

Beyond the GW method: Simple vertex correction improves band energies of bulk and two-dimensional materials
Per S. Schmidt, Christopher E. Patrick and Kristian S. Thygesen
Submitted

Here different ways of including non-trivial vertex corrections in quasiparticle calculations is discussed and the effect this has on the resulting band gaps and band gap centres is explained. In particular, the focus is on the approach presented in the attached publication where many-body perturbation theory is combined with time-dependent density functional theory to introduce a vertex correction which is designed to improve short-range electron correlations. With this, improved accuracy of absolute band positions as well as band gaps is achieved and the long-standing theoretical problem regarding the inconsistency of GW calculations with their standard starting point is resolved. Furthermore, the method is shown to have improved convergence behaviour, meaning that higher accuracy can be achieved at a lower computational cost.

4.3.1 Iterating Hedin's equations

The most obvious way to go beyond the GW approximation and include a non-trivial vertex function is to perform another iteration of Hedin's equations starting from $\Sigma = iGW$. Neglecting derivatives of W produces the kernel $K(1,2,3,4) = \delta\Sigma(1,2)/\delta G(3,4) = iW(1,2,3,4)$, which is known from the Bethe-Salpeter Equation. The four-point nature of this kernel makes it difficult to invert the vertex equation, $\Gamma = \delta + KGG\Gamma$, without loss of accuracy. Instead one can perform a truncation of the Dyson equation for the vertex to obtain $\Gamma = \delta + WGG$, which leads to a self-energy consisting of a second-order screened exchange term in addition to the usual iGW term. This method is termed GWΓ^1 and Gruneis et al. have shown[98], using a static approximation for W in the vertex, that this GWΓ^1 approximation, performed in a fully self-consistent manner, leads to significant improvements for band gaps and ionization potentials of solids. From a theoretical point of view this is a highly satisfactory result. The drawback is the higher complexity of the formalism and the concomitant loss of physical transparency as well as the significant computational overhead as compared to the GW method.

Time-dependent density-functional theory (TDDFT) in principle offers a framework for including exchange-correlation (xc)-effects in the dynamical response via a two-point vertex function rather than the computationally challenging three-point vertex function that arises naturally in the diagrammatic many-body formalism. While it appears attractive to use TDDFT derived vertex functions in many-body calculations, progress along these lines has been hindered by the poor quality of the local xc-kernels based on standard local xc-potentials. However, recent work has shown that a simple renormalization of the adiabatic LDA xc-kernel can overcome these problems and yield a dramatic improvement over the RPA for total energy calculations based on the adiabatic connection fluctuation dissipation theorem (ACDFT), as presented in section 3.2.

As shown in Chapter 4, iterating Hedin's equation starting from $\Sigma = 0$ (the Hartree approximation) leads to the standard GW expression which in turn should be evaluated on Hartree wave functions and energies. In the following, the iterations are carried out from the starting point of the self-energy $\Sigma(1,2) = \delta(1,2)v_{xc}^{\mathrm{LDA}}(1)$ [88, 99]. This justifies the use of LDA input wave functions and energies.
The density can be written in terms of the Green's function, $n(1) = -iG(1,1^+)$, and the vertex then becomes

$$\Gamma(1,2,3) = \delta(1,2)\delta(1,3) - i\delta(1,2)\int d(5,6,7)\frac{\delta v_{xc}^{\mathrm{LDA}}(1)}{\delta n(5)}G(5,6)G(7,5)\Gamma(6,7,3).$$

Defining the kernel as $f_{xc}(1,5) = \frac{\delta v_{xc}^{\mathrm{LDA}}(1)}{\delta n(5)}$ and introducing a new function

$$k(1,3) = \delta(1,3) + \int d(5,6)f_{xc}(1,5)\chi_0(5,6)k(6,3) \tag{4.11}$$

with the usual definition of the single-particle polarizability $\chi_0(1,2) = -iG(1,2)G(2,1)$, it is clear that we can factorize the vertex function as

$$\Gamma(1,2,3) = \delta(1,2)k(1,3).$$

Inserting this into Eq. (4.4) gives

$$P(1,2) = \int d3\chi_0(2,3)k(3,1).$$

For ease of notation, the formal solution of Eq. (4.11) can be written in operator form as

$$k = (1 - f_{xc}\chi_0)^{-1}.$$

Similarly, Eq. (4.5) solves to

$$W = v(1 - v\chi_0 k)^{-1}.$$

Inserting into Eq.(4.3) yields

$$\begin{aligned}\Sigma &= iGv(1 - v\chi_0 k)^{-1}(1 - f_{xc}\chi_0)^{-1} \\ &= iGv(1 - (v + f_{xc})\chi_0)^{-1} \\ &= iG\widetilde{W}\end{aligned}$$

which takes the same form as the standard GW method but with a different screened Coulomb interaction. This method we term GWΓ.
If we set $\Gamma(1,2,3) = \delta(1,2)\delta(2,3)$ in Eq. (4.3) but use $\Gamma(1,2,3) = \delta(1,2)k(1,3)$ in Eq. (4.4) we instead arrive at

$$\begin{aligned}\Sigma = iGv(1 - v\chi_0 k)^{-1} &= iGv\Big(1 - v\chi_0(1 - f_{xc}\chi_0)^{-1}\Big)^{-1} \\ &= iGv\Big(1 - f_{xc}\chi_0\Big)\Big(1 - (v + f_{xc})\chi_0\Big)^{-1} \\ &= iGv\Big(1 + v\chi_0\Big[1 - (v + f_{xc})\chi_0\Big]^{-1}\Big) \\ &= iG\widehat{W}\end{aligned}$$

which we term the GWP method. There is no apparent physical justification for this choice but it has been shown a priori to give reasonable results when combined with self-consistency[100].
For completeness, there is of course the option of setting $\Gamma(1,2,3) = \delta(1,2)\delta(2,3)$

in Eq. (4.4) but keeping the full expression in Eq. (4.3), resulting in:

$$
\begin{aligned}
\Gamma &= 1 + f_{xc}\chi_0\Gamma = 1 + f_{xc}\chi_0 + f_{xc}\chi_0 f_{xc}\chi_0 + \cdots \\
W &= v + vPW = v + vPv + vPvPv + \cdots \\
P &= \chi_0 \\
&\Downarrow \\
\Sigma &= iGW\Gamma = iG(v + v\chi_0 v + v\chi_0 v\chi_0 v + \cdots)(1 + f_{xc}\chi_0 + f_{xc}\chi_0 f_{xc}\chi_0 + \cdots) \\
&= iG(v + v(v + f_{xc})\chi_0 + v\chi_0(v + f_{xc})v\chi_0 + \cdots) \\
&= iG(v + v\chi(v + f_{xc})) \\
&= iG(v + v\chi_0(1 - v\chi_0)^{-1}(v + f_{xc})) \\
&= iG\overline{W}
\end{aligned}
$$

This method we term $GW\Sigma$.
It is worth noting that the same four expressions can be obtained by writing

$$
\begin{aligned}
\Sigma &= iG(v + v\chi(v + f_{xc})) \\
\chi &= \chi_0 + \chi_0(v + f_{xc})\chi
\end{aligned}
$$

with $f_{xc} \in [0, \mathrm{rALDA}]$. Including f_{xc} in the expression for Σ implies that xc effects are included in the potential created by an induced density. Including f_{xc} in χ reduces the screening in the system.
The size of the band gap and the absolute positions of the ionization potential and electron affinity of bulk BN is calculated with the four methods and shown in Fig. 4.3. It is evident that two different effects occur, depending on where the xc-kernel is included. Both the size of the band gap and the band gap center is affected. The reason for the shifts are explained in the following section and the attached publication. The explanation requires that we first establish a connection between the quasiparticle energies obtained via the GW framework and total energies from the RPA.
The four methods are summarized in Table 4.1.

	Σ	χ	iGW
GW	RPA	RPA	$iGv(1-\chi_0 v)^{-1}$
GWP	RPA	rALDA	$iGv(1-\chi_0 f_{xc})(1-\chi_0(v+f_{xc}))^{-1}$
GWΓ	rALDA	rALDA	$iGv(1-\chi_0(v+f_{xc}))^{-1}$
GWΣ	rALDA	RPA	$iGv(1+\chi_0(1-v\chi_0)^{-1}(v+f_{xc}))$

Table 4.1: Summary of the various quasiparticle methods.

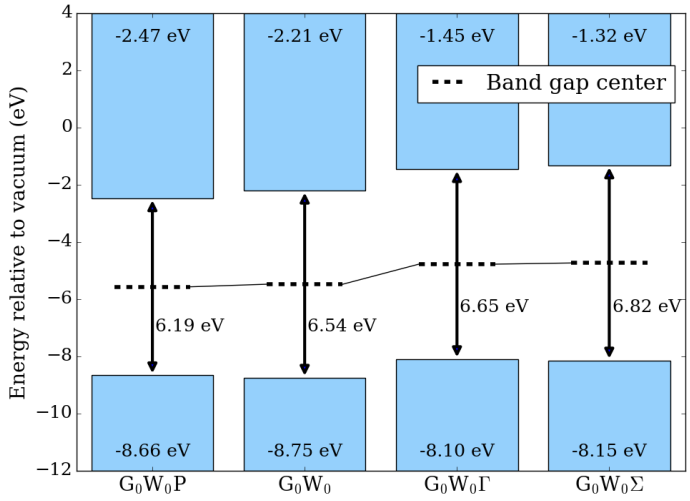


Figure 4.3: CBM and VBM of bulk BN relative to vacuum calculated with four different methods (see text). For computational details, see the attached publication.

4.3.2 Quasiparticle energies from total energy calculations

In the following section physical meaning is given to the shift in band gaps and band gap centers obtained for BN with the four approximations of the previous section. To do this, the quasiparticle energies are first approached from a total energy perspective.

The ionization potential (IP) and electron affinity (EA) are defined from total energy differences between N- and $N \pm 1$-electron systems

$$\begin{aligned} \text{IP} &= E[n_{N-1,N-1}] - E[n_{N,N}] \simeq E[n_{N-1,N}] - E[n_{N,N}] \\ \text{EA} &= E[n_{N,N}] - E[n_{N+1,N+1}] \simeq E[n_{N,N}] - E[n_{N+1,N}] \end{aligned}$$

where $n_{N,M}$ is the N-electron density constructed by the N lowest KS orbitals of the M-electron system. The $\simeq$ becomes an equality assuming no orbital relaxation. With this definition the IP and EA become positive quantities. In turn, the absolute position of the valence (conduction) band is the negative of the IP (EA). Neglecting correlation, the EXX IP and EA are obtained as shown in green in Fig. 4.4; the IP is at -10.65 eV and the EA at 2.56 eV. EXX clearly overestimates the gap and predicts the IP (EA) to be positioned

much lower (higher) than when including correlation via the G_0W_0 method shown in blue. In the following it is first explained why adding the correlation contribution leads to a closing of the gap.

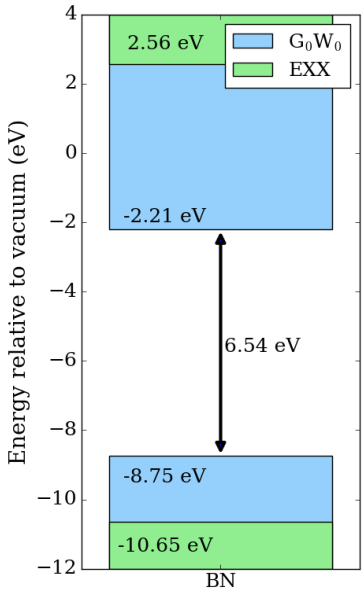


Figure 4.4: EXX and G_0W_0 calculations of the IP and EA of bulk BN.

Niquet et al.[101] has shown how the correlation part of the total energy differences calculated within RPA is directly related to the correlation part of the quasiparticles from a GW calculation. The RPA correlation energy is calculated from the non-interacting density response function:

$$E_c^{\mathrm{RPA}}[n] = \frac{1}{2}\int \frac{d\omega}{2\pi}\mathrm{Tr}\Big\{\ln[1-v\chi_0(i\omega)]+v\chi_0(i\omega)\Big\}$$

from which it is clear that the differences in total energies between the N- and $N \pm 1$-electron systems come from a change in χ_0: $\chi_0 \to \chi_0 + \delta\chi_0$. The resulting contribution from the correlation energy to the absolute position of the conduction band is

$$\begin{aligned} -\mathrm{EA}_c^{\mathrm{RPA}} &= E_c^{\mathrm{RPA}}[n_{N+1,N}] - E_c^{\mathrm{RPA}}[n_{N,N}] \\ &= -\frac{1}{2}\int \frac{d\omega}{2\pi}\mathrm{Tr}\Big\{[W(i\omega)-v]\delta\chi_0(i\omega)\Big\}, \end{aligned}$$

where $W(i\omega) = v[1-v\chi_0(i\omega)]^{-1}$. The change in $\chi_0(i\omega)$ when adding one electron to the lowest unoccupied KS orbital, $\psi_c(\mathbf{r})$, can be written in terms of the Green's function,

$$\delta\chi_0(\mathbf{r},\mathbf{r}',i\omega) = \psi_c^*(\mathbf{r})G_0(\mathbf{r},\mathbf{r}',\epsilon_c+i\omega)\psi_c(\mathbf{r}') + \text{c.c.},$$

which makes it possible to write the correlation contribution to the EA in terms of the matrix element of the GW correlation-self-energy with the conduction band:

$$-\mathrm{EA}_c = \langle\psi_c|\Sigma_c(\epsilon_c)|\psi_c\rangle$$

A similar expression can be obtained for the IP. Thus, in order to explain the down-shift (up-shift) of the conduction (valence) band when including correlation (from EXX to G_0W_0), it is enough to show that the absolute correlation energy of the $N \pm 1$-electron system is more negative than that of the N-electron system.

χ_0 consists of contributions between occupied and unoccupied states which in real space can be written

$$\chi_0(\mathbf{r},\mathbf{r}',\omega) = -\sum_n^{\text{occ.}}\sum_m^{\text{unocc.}} 2\left(\frac{\psi_m^*(\mathbf{r}')\psi_n(\mathbf{r}')\psi_n^*(\mathbf{r})\psi_m(\mathbf{r})}{\epsilon_m-\epsilon_n-\omega}+\frac{\psi_n^*(\mathbf{r}')\psi_m(\mathbf{r}')\psi_m^*(\mathbf{r})\psi_n(\mathbf{r})}{\epsilon_m-\epsilon_n+\omega}\right).$$

Compared with the N-electron system, in the $N+1$-electron system there will be additional contributions from all unoccupied states to the now occupied state in the conduction band. This picture is sketched in Fig. 4.6. There will no longer be contributions from the same state with all the occupied ones:

$$\begin{aligned}\delta\chi_0 = &-\sum_n^{\text{occ.}} 2\left(\frac{\psi_c^*(\mathbf{r}')\psi_n(\mathbf{r}')\psi_n^*(\mathbf{r})\psi_c(\mathbf{r})}{\epsilon_c-\epsilon_n-\omega}+\frac{\psi_n^*(\mathbf{r}')\psi_c(\mathbf{r}')\psi_c^*(\mathbf{r})\psi_n(\mathbf{r})}{\epsilon_c-\epsilon_n+\omega}\right)\\ &+\sum_m^{\text{unocc.}} 2\left(\frac{\psi_m^*(\mathbf{r}')\psi_c(\mathbf{r}')\psi_c^*(\mathbf{r})\psi_m(\mathbf{r})}{\epsilon_m-\epsilon_c-\omega}+\frac{\psi_c^*(\mathbf{r}')\psi_m(\mathbf{r}')\psi_m^*(\mathbf{r})\psi_c(\mathbf{r})}{\epsilon_m-\epsilon_c+\omega}\right)\end{aligned}$$

The two terms come with different sign and will as such partly cancel each other out. In large-gap insulators, the first term will be negligible as $\epsilon_c-\epsilon_n$ is large. In the second term, $\epsilon_m > \epsilon_c$ making $\delta\chi_0 > 0$ and hence $\chi_0[N] > \chi_0[N+1]$. A similar argument results in $\chi_0[N] > \chi_0[N-1]$.

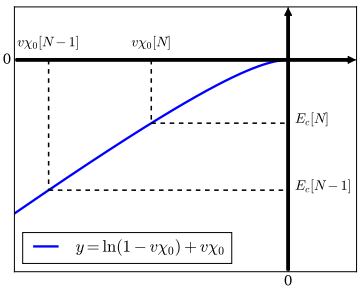


Figure 4.5: Sketch of the integrand in the RPA correlation energy.

The correlation energy, $E_c^{\text{RPA}} \propto \ln(1-v\chi_0)-v\chi_0$ is shown in Fig.4.5 as a function of $v\chi_0$ and it is evident from $\chi_0[N] > \chi_0[N\pm1]$ that $E[N] > E[N\pm1]$. This means that the correlation contribution

$$\begin{aligned}-\text{IP}_c^{\text{RPA}} &= E_c^{\text{RPA}}[n_{N,N}] - E_c^{\text{RPA}}[n_{N-1,N}] > 0\\ -\text{EA}_c^{\text{RPA}} &= E_c^{\text{RPA}}[n_{N+1,N}] - E_c^{\text{RPA}}[n_{N,N}] < 0\end{aligned}$$

which explains the downward (upward) shift of the conduction (valence) bands going from EXX to G_0W_0. This picture is a simplification but it provides the right trends. An explanation as to why the downward shift of the conduction band is much larger than the upward shift of the valence band, we have split the correlation contribution in two terms; the correlation energy of an atom in the ground state and a stabilising screening contribution. The discussion and implications of this can be found in the attached publication.

In the following, the change in the size of the band gap and absolute band positions in Fig. 4.3 when applying the four different methods is briefly addressed. Further discussion is found in the attached publication.

From Fig. 4.3 we observe that the dominant effect of the vertex corrections is to shift the band gap center while the band gap itself is less affected. Physically, the main effect of the rALDA kernel is to modify the effective Coulomb interaction at short distances, but it also reduces the long range components slightly. This suggests that the band gap is mainly determined by long range interactions while the band gap center is sensitive to the short range correlations.

Including f_{xc} in χ (GWP and GWΓ) affects the description of screening while including it in Σ (GWΓ and GWΣ) affects the form of the potential created by the induced density. From Fig. 4.3 it is clear that the band gap center depends mainly on f_{xc} in the self-energy, i.e. a correct description of the band gap center requires the inclusion of xc-effects in the induced potential. The band gaps increase in the order GWP, GW, GWΓ, GWΣ. The size of the gap depends mainly on the long range interactions and f_{xc} reduces these. Thus the total induced potential will generally be smaller when xc-effects are included in the final potential (GWΓ and GWΣ). More precisely, given a density variation, δn, the corresponding induced electron potential, $\delta v_{\mathrm{Hxc}} = (v + f_{xc})\delta n$, is generally weaker than the bare Hartree potential $\delta v_H = v\delta n$, because $v(q) + f_{xc}(q) < v(q)$. This explains the larger band gaps found for the GWΓ and GWΣ methods. The remaining ordering (GW $>$ GWP and GWΣ $>$ GWΓ) comes from noting that $\chi^{\mathrm{rALDA}} > \chi^{\mathrm{RPA}}$ because the higher order diagrams, which reduce the effect of χ_0, are larger in RPA

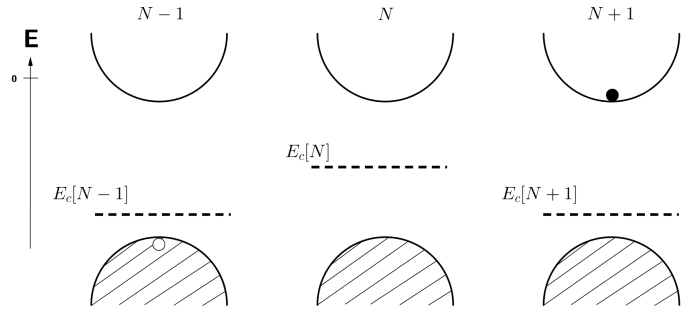


Figure 4.6: Sketch of the difference in correlation energy between the $N \pm 1$- and the N-electron system.

because $v(q) + f_{xc}(q) < v(q)$.

In the attached publication, three of the methods described above are tested on a wider range of semi-conductors, both 2D and 3D systems. The best agreement with experimental band gaps and absolute band positions is obtained for the GWΓ method, which at the same time shows improved convergence behaviour. The methods come with practically the same computational cost, hence the GWΓ provides an efficient and rigorous improvement over the GW approximation and sets a new standard for quasiparticle calculations in solids.

CHAPTER 5

Connecting GW with DFT: Optimized Effective Potential

In the previous Chapter 4 it was shown how to calculate excitations in solids using MBPT and the QP picture. In this chapter we return to the Kohn-Sham formalism and discuss the physics behind the eigenvalues obtained when applying the RPA to derive a local xc potential. This requires linking MBPT to DFT through the Sham-Schlüter equation[102]. The method, also called the optimized effective potential (OEP) method, addresses solving the Sham-Schlüter equation to obtain the local RPA xc potential from the non-local GW self-energy. The xc potential can in turn be used directly in the Kohn-Sham equations. From solving the Kohn-Sham equations one obtains the resulting RPA equivalent of KS eigenvalues, $\epsilon_i^{\mathrm{RPA}}$, and wave functions, $\psi_i^{\mathrm{RPA}}(\mathbf{r})$. Due to the incorporation of exact exchange, these wave functions are therefore likely not to suffer from delocalization errors, that are known to be present in LDA or GGA wave functions.
As mentioned previously, one issue with the standard RPA method is the lack of self-consistency. By evaluating the RPA total energy expression using $\epsilon_i^{\mathrm{RPA}}$ and $\psi_i^{\mathrm{RPA}}(\mathbf{r})$ one obtains the true self-consistent RPA total energy.

The fist part of this chapter discusses self-consistent RPA total energy calculations and the specific implementation. The second part addresses the Kohn-Sham eigenvalues obtained when using the RPA to the xc potential in the OEP method.

5.1 Self-consistent RPA calculations

The RPA correlation energy depends on the orbitals for which it is evaluated. This is exemplified in Fig. 5.1 (from [54]) where the RPA atomization energy for simple molecules is seen to vary with up to 10 kcal/mol depending on whether LDA or PBE orbitals and energies were used as input. The obvious question is then what is the true RPA value if evaluated on RPA wave func-

tions and energies? Whether errors in e.g. atomization energies, compared

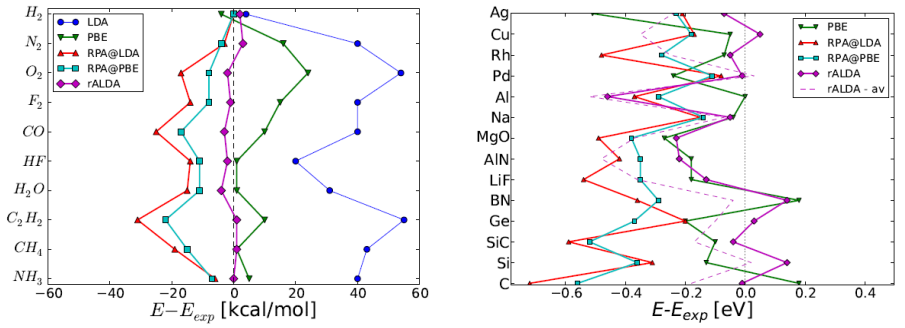


Figure 5.1: **(a)** Atomization and **(b)** cohesive energies calculated within RPA using either PBE or LDA input wave functions, from [54].

with experiments, are related to the used orbitals or are intrinsic to the RPA can only be answered by carrying out a self-consistent treatment of the RPA (scRPA)[57]. In scRPA, the wave functions used to evaluate $E_{\mathrm{EXX}}[n]$ and $E_c^{\mathrm{RPA}}[n]$ are those from a self-consistent DFT ground state calculation with $v_{\mathrm{xc}} = v_{\mathrm{xc}}^{\mathrm{RPA}}$. The rest of this section addresses how to obtain $v_{\mathrm{xc}}^{\mathrm{RPA}}$, and from that the RPA wave functions and energies necessary to evaluate $E_{\mathrm{EXX}}[n]$ and $E_c^{\mathrm{RPA}}[n]$. The method is outlined in the following and a sketch is provided in Fig. 5.2.

- The ground state (e.g. LDA) wave functions are used to calculate the first iteration of $v_{\mathrm{xc}}^{\mathrm{RPA}}$.
- $v_{\mathrm{xc}}^{\mathrm{LDA}}$ is replaced with $v_{\mathrm{xc}}^{\mathrm{RPA}}$ in the Kohn-Sham scheme and the Kohn-Sham equation is solved using this new xc potential.
- The resulting wave functions are used to calculate the second iteration of $v_{\mathrm{xc}}^{\mathrm{RPA}}$ and the loop continues until self-consistency in the potential is reached.
- The ground state energy is then evaluated using the resulting RPA ground state wave functions.

The RPA method is present in various electronic structure codes and the idea of scRPA has therefore also been present in the community for a long time. The opinion on what there is to gain from such a calculation has been divided. Some believe a self-consistent approach is necessary:

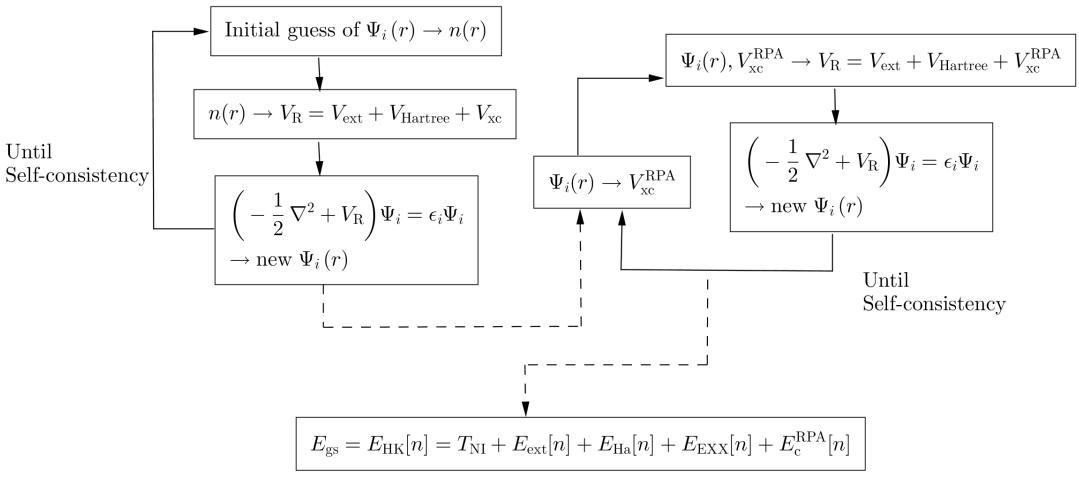


Figure 5.2: Scheme for scRPA groundstate energy calculations. Solid lines indicate self-consistency loops, dashed lines indicate that the output of the loops is used onwards in the calculations.

> "*Our own work indicates that the input orbital dependence in RPA post-processing calculations is a significant issue. Some form of self-consistency would therefore be desirable.*" (Ren, Rinke, Joas, Scheffler 2012[49])

Some do not expect much to be gained for bulk systems:

> "*For solids, calculations presented so far suggests that the RPA-KS potential exhibits only small differences to the LDA and PBE potentials, suggesting that self-consistency is a minor issue for extended systems.*" (Harl, Schimka, Kresse 2010[57])

Similar for atomic systems:

> "*Our experience with conventional density functionals suggests that, at least for atomic systems, the RPA correlation energy is not sensitive to the detailed structure of the exchange-correlation potential.*" (Jiang, Engel 2007[61]),

but in the end:

> "*Only a self-consistent treatment of the RPA will allow a final answer to whether errors in atomization energies are related to the*

> *used orbitals or are intrinsic to the RPA.*" (Harl, Schimka, Kresse 2010[57])

The fundamental point is that

> "*A self-consistent implementation is only advantageous if the resulting correlation potential leads to an improved total Kohn-Sham potential.*" (Jiang, Engel 2007[61])

For this method to be useful, one must examine a quantity which is sensitive to the exact shape of the correlation potential. According to [57] lattice constants and bulk moduli are not. For systems such as transition metal oxides (TMO) and nitrides it is known that GGA orbitals are too delocalized due to an overestimation of the energy of the localized d-states. It is expected that RPA wave functions will not suffer from delocalization to the same degree since the self-interaction errors are mostly removed through the incorporation of exact exchange. Formation energies of the TMOs and adsorption energies on selected TMO surfaces could therefore be areas where results would be improved when employing self-consistency.

Additionally, having access to the RPA xc-potential would enable the efficient evaluation of atomic forces at the RPA level, accounting for long-range interactions, for which self-consistency is a fundamental requirement. This is key in advancing the RPA method by allowing structural relaxations of complex systems and molecular dynamics. The evaluation of RPA forces is however beyond the scope of this chapter.

The RPA xc-potential is obtained from the linearized Sham-Schlüter equation which relates the local RPA xc-potential to the non-local GW many-body self-energy. This will be presented in greater detail in the following section.

5.2 The Sham-Schlüter equation

The total energy expressions for the EXX and RPA correlation energies were presented in Eq.(2.2) and (3.4) respectively. The corresponding xc-potential can then by definition be obtained directly from $v_{xc}^{\mathrm{RPA}} = \frac{\delta E_{xc}^{\mathrm{RPA}}}{\delta n}$. As both the EXX and RPA correlation expressions are explicit functionals of the ground state wave functions and hence implicit functionals of the density, carrying out the functional derivative is somewhat extensive. The details can be found in [103]. The result is known as the linearized Sham-Schlüter equation and is

repeated here with the GW approximation to the self-energy:

$$\int d\mathbf{r}_3 v_{\mathrm{xc}}^{\mathrm{RPA}}(\mathbf{r}_3) \int \frac{d\omega}{2\pi} G_0(\mathbf{r}_1, \mathbf{r}_3, i\omega) G_0(\mathbf{r}_3, \mathbf{r}_1, i\omega)$$
$$= \int \frac{d\omega}{2\pi} \int d\mathbf{r}_3 d\mathbf{r}_4 G_0(\mathbf{r}_1, \mathbf{r}_3, i\omega) \Sigma_{xc}^{\mathrm{GW}}(\mathbf{r}_3, \mathbf{r}_4, i\omega) G_0(\mathbf{r}_4, \mathbf{r}_1, i\omega)$$

with the GW self-energy, as presented earlier,

$$\Sigma_{xc}^{\mathrm{GW}}(\mathbf{r}_1, \mathbf{r}_2, i\omega) = \frac{i}{2\pi} \int d\omega' G_0(\mathbf{r}_1, \mathbf{r}_2', i\omega + i\omega') W(\mathbf{r}_1, \mathbf{r}_2', i\omega')$$

The Sham-Schlüter equation can also be obtained using the property that the density of the interacting and reference KS system are the same[102].
The equation simplifies when only including the exchange part of the self-energy as that is frequency independent[104]. This allows us to carry out the frequency-integration of the right hand side using contour integration and arrive at

$$\Sigma_x(\mathbf{r}_1, \mathbf{r}_2) = \sum_n^{\text{occ.}} \frac{\psi_n(\mathbf{r}_1)\psi_n^*(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \Rightarrow$$
$$\int d\mathbf{r}_3 v_{\mathrm{x}}^{\mathrm{EXX}}(\mathbf{r}_3)\chi_0(\mathbf{r}_3, \mathbf{r}) = \sum_i^{\text{occ.}} \sum_a^{\text{unocc.}} \int d\mathbf{r}_3 \int d\mathbf{r}_4 \frac{1}{\epsilon_i - \epsilon_a} \psi_a(\mathbf{r}_1)\psi_a^*(\mathbf{r}_3)$$
$$\times \Sigma_x(\mathbf{r}_3, \mathbf{r}_4)\psi_i(\mathbf{r}_4)\psi_i^*(\mathbf{r}_1) + \text{c.c.}$$

where the frequency integration of the Green's functions on the LHS

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{1}{i\omega - \epsilon_{n\mathbf{k}}} \frac{1}{i\omega - \epsilon_{n'\mathbf{k}'}} = \frac{f_{n\mathbf{k}} - f_{n'\mathbf{k}'}}{\epsilon_{n\mathbf{k}} - \epsilon_{n'\mathbf{k}'}}$$

with the occupation numbers f_{nk}. This makes it possible to write the LHS in terms of the static non-interacting response function. We refer to this as the EXX-OEP equation.

With the full GW self-energy, the frequency dependence makes the solution of the OEP computationally more involved. To simplify the calculations, the full frequency dependence of the self-energy is approximated by the self-energy at the poles. This is also known as the quasi-particle GW approximation (qpGW):

$$\langle n\mathbf{k}|\Sigma_{xc}^{\mathrm{qpGW}}|m\mathbf{k}\rangle \equiv \Sigma_{nm}^{qpGW} \approx \frac{1}{2} \langle n\mathbf{k}|[\Sigma_{xc}^{\mathrm{GW}}(\epsilon_{n\mathbf{k}})]^* + \Sigma_{xc}^{\mathrm{GW}}(\epsilon_{m\mathbf{k}})|m\mathbf{k}\rangle$$

where the complex conjugation in the first term makes the self-energy Hermitian. Removing the frequency dependence of the self-energy means that the frequency-integration of the Green's functions can be carried out analytically, as was the case for the EXX-OEP. The approximate RPA-OEP equation thus becomes similar to the EXX-OEP one, except the matrix elements, $\langle n|\Sigma_x|m\rangle$ need to be replaced with Σ_{nm}^{qpGW}. Σ_{nm}^{qpGW} is evaluated in a plane wave basis as in a standard GW calculation, see Eq.(4.9). The only difference is that off-diagonal matrix elements are required, $n \neq m$. This does however increase the computational time significantly. Fourier-transforming the entire RPA-OEP results in

$$v_{xc}^{\mathrm{RPA}}(\mathbf{G}) = \sum_{\mathbf{G}'} \chi_0^{-1}(\mathbf{G},\mathbf{G}') \sum_{n\mathbf{k},m\mathbf{k}} \langle n\mathbf{k}|e^{i\mathbf{G}'\cdot\mathbf{r}}|m\mathbf{k}\rangle \frac{\Sigma_{nm}^{\mathrm{qpGW}}}{\epsilon_{m\mathbf{k}} - \epsilon_{n\mathbf{k}}} + \text{c.c.} \qquad (5.1)$$

and v_{xc}^{RPA} can then be obtained in real-space by the inverse Fourier transform

$$v_{xc}^{\mathrm{RPA}}(\mathbf{r}) = \frac{1}{N}\sum_{\mathbf{G}} v_{xc}^{\mathrm{RPA}}(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}}$$

where N is the number of plane waves.

5.2.1 PAW corrections

The terms $\chi_0^{-1}(\mathbf{G},\mathbf{G}')$, $\langle n\mathbf{k}|e^{i\mathbf{G}'\cdot\mathbf{r}}|m\mathbf{k}\rangle$ and $\Sigma_{nm}^{\mathrm{qpGW}}$ in Eq.(5.1) are all constructed from wave functions which consist of a pseudo-part plus PAW corrections, as discussed in Section 2.4.1. In order to obtain the all-electron RPA xc potential one should include the pseudo wave functions plus the PAW corrections in the evaluations of the expressions above.

When applying the RPA xc potential in the Kohn-Sham formalism to obtain the RPA eigenvalues and wave functions, it is however not the all-electron xc potential that enters the equations. In Section 2.4.1 it was shown how the PAW-transformed Hamiltonian takes the form

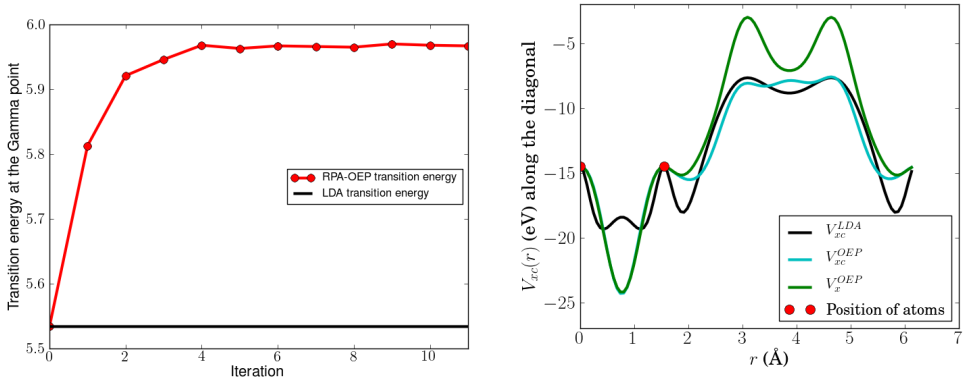
$$\tilde{\hat{H}} = -\frac{1}{2}\nabla^2 + v_{\mathrm{Ha}}[\tilde{n}] + v_{xc}[\tilde{n}] + \sum_a \sum_{i_1 i_2} |\tilde{p}_{i_1}^a\rangle \, \Delta H_{i_1 i_2}^a \, \langle \tilde{p}_{i_2}^a| \,.$$

We thus need the RPA xc potential evaluated on the pseudo density, $v_{xc}^{\mathrm{RPA}}[\tilde{n}]$ to replace $v_{xc}^{\mathrm{LDA}}[\tilde{n}]$, and the PAW corrections must enter $H_{i_1 i_2}^a$ through carrying out the differentiation $\frac{\delta \Delta E_{xc}^{\mathrm{RPA}}}{\delta D_{i_1 i_2}^a}$. It was not possible to carry out this differentiation analytically for the RPA correlation energy and as an approximation, this implementation keeps the PAW corrections from the LDA whilst

replacing the pseudo LDA xc potential with the RPA counterpart. In [104] the PAW corrections are obtained from a linearized Hartree-Fock method and kept fixed for EXX-OEP and RPA-OEP calculations.

5.3 Transition energies

The implementation of the OEP equation and introducing $v_{xc}^{\mathrm{RPA}}(\mathbf{r})$ in the KS loop in GPAW alongside with PAW corrections and parallellisation schemes was the focus of this project initially. The method was thought to be applied to transition metal oxides (TMOs) where the complex electronic structure arising from the localized and strongly correlated *d*-electrons has previously made theoretical descriptions highly challenging [105, 106, 107]. TMOs are of great importance for a number of technological applications such as transparent electrodes, solid electrolytes, superconductors, microelectronics, computer chips, batteries and piezoelectrics. Self-consistent RPA calculations was thought of as a way of dealing with the delocalization errors otherwise present in DFT with standard xc functionals. This could in turn be used to construct a comprehensive computational database of TMO surface chemistry properties that will fill the "data gap" existing for this important class of materials. This project was however abandoned due to reasons explained in the following sections.



(a) (b)

Figure 5.3: Diamond: **(a)** The $\Gamma \to \Gamma$ transition energy of diamond as a function of iterations in the self-consistent cycle. **(b)** Converged xc OEP potential, exchange-only OEP potential ($E_{\mathrm{cut}} = 300\,\mathrm{eV}$, k-pts= $6 \times 6 \times 6$) and the LDA xc potential (aligned at $\mathbf{r} = 0$).

With the implementation available in GPAW, we decided to start off by looking at band gaps of various semi-conductors, calculated from the difference between the lowest unoccupied and highest occupied Kohn-Sham eigen-

value, to compare our results with recently published results from a similar implementation in the VASP code[104]. Even given the exact Kohn-Sham xc-potential, the resulting band gap will by construction not be exact because of the lack of the derivative discontinuity, as discussed in Section 2.3.1. As shown in [36, 104] the derivative discontinuity can be obtained from a G_0W_0 calculation but that is neglected here. It is only of interest to see how the Kohn-Sham formalism performs when employing the RPA-OEP xc-potential.

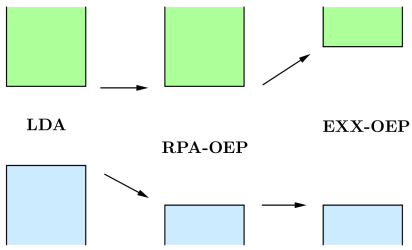


Figure 5.4: Sketch of the shift in energy levels. Blue (green) denotes valence (conduction) states.

A simple example is diamond with only two atoms in the unit cell and a rather small lattice constant. In Fig. 5.3(a) the band gap at Γ is shown as a function of the number of iterations of the RPA-OEP equation (second loop of Fig. 5.2). This calculation is carried out at a plane wave cutoff of 300 eV and a $6 \times 6 \times 6$ k-points grid. The number of bands included in the construction of $\Sigma_{nm}^{\mathrm{qpGW}}$ was set equal to the number of plane waves. The band gap is seen to be converged after approximately 5 iterations. Throughout the iterations, 70 % of the newly calculated RPA xc-potential is mixed with 30% of the old to ensure smoother convergence.
The numbers for the band gap of diamond with both EXX- and RPA-OEP are printed in Table. 5.1 for plane wave cutoffs of 300 and 600 eV as well as after an extrapolation to infinity. The results from the VASP code are reprinted as well, showing excellent agreement.

In order to explain the relative size differences between the LDA, EXX-OEP and RPA-OEP band gap, it is necessary to look at the RPA xc-potential in greater detail. The resulting RPA xc-potential for diamond is shown in Fig. 5.3(b) plotted through the diagonal of the unitcell, hence going through the

	LDA		EXX-OEP				RPA-OEP				Exp.
Code:	GPAW	VASP	GPAW			VASP	GPAW			VASP	-
$E_{\mathrm{cut}}^{\mathrm{PW}}$ (eV):	800	-	300	600	∞	∞	300	600	∞	∞	-
diamond	5.53	5.54	6.17	6.18	6.19	6.20	5.97	6.02	6.05	6.07	7.30

Table 5.1

two carbon atoms at $r = 0$ and $r \approx 1.7$ Å. It is evident from Tab. 5.1 that LDA predicts the smallest gap followed by RPA-OEP and EXX-OEP. The EXX-OEP gap is in decent agreement with experimental values. EXX-OEP gaps for five different semi-conductors are shown in Table 5.2 from which it is clear that the agreement for small gaps can only be accidental as it fails completely for larger gap systems such as Ne. There is no physical reason why EXX-OEP gaps should be similar to experimental quasiparticle gaps. The agreement here is attributed to error cancellation between the lack of correlation and omitting the derivative discontinuity[104].
From Fig. 5.3b we see that the EXX-OEP potential is more attractive than the LDA potential in the bonding region but more repulsive in the interstitial regions. Compared with the EXX-OEP potential, the LDA potential is more spherically symmetric around the atoms (red dots). This is attributed to the one-electron self-interaction free nature of the EXX-OEP. The RPA-OEP potential is also deeper in the bonding region, being almost identical to the EXX-OEP potential. However, the RPA-OEP potential differs significantly from the EXX-OEP one in the interstitial region, where it much resembles the LDA potential. The RPA-OEP potential varies very little in the interstitial regions, especially compared to the EXX-OEP potential. This can be attributed to the fact that the method incorporates the response of the surrounding electron density (screening), not included in EXX-OEP[104].
The LDA and RPA-OEP potentials are similar in the interstitial regions of low density, but the RPA-OEP potential, is more attractive in the bonding region where the density of the valence band states is large. This lowers the position of the valence band states and increases the gap compared to LDA. The EXX-OEP and RPA-OEP potentials are similar in the bonding region, resulting in similar positions of the valence band edges. The differences occur in the interstitial regions where the RPA-OEP potential is more attractive. As the interstitial regions are dominated by the conduction band states, the more repulsive EXX-OEP potential increases the energy of the conduction band states so that the EXX-OEP gap becomes even wider than the RPA-OEP gap. This is sketched in Fig.5.4. In conclusion, the resulting numbers and observed trends are in excellent agreement with that of J. Klimes et al. [104], supporting the implementation made in GPAW.

	LDA		EXX-OEP				Exp.
Code:	GPAW	VASP	GPAW			VASP	-
PW Cutoff (eV):	800	-	300	600	∞	∞	-
Si	2.53	2.53	2.90	2.91	2.92	3.13	3.05
BN	8.69	8.69	9.95	9.99	10.01	9.84	-
GaN	1.70	1.69	3.60	3.47	3.40	3.12	3.30
MgO	4.68	4.68	7.60	7.39	7.28	6.64	7.83
Ne	11.41	11.43	14.89	14.59	14.43	14.69	21.51

Table 5.2: EXX-OEP results for a range of semi-conductors compared with that of J. Klimes et al.[104].

5.4 The potential of the method and computational cost

In [104] the derivative discontinuity is added to the RPA-OEP gaps to obtain the quasiparticle gap at Γ of 8.00 eV. This results in an overestimation compared with experiments. The derivative discontinuity is obtained from a G_0W_0@LDA calculation, which by itself yields seemingly better agreement. With GPAW, the band gap of diamond at Γ is 7.56 eV calculated with G_0W_0@LDA, which comes at a much smaller computational cost and is closer to the experimental value of 7.30 eV.

This is hardly satisfactory as the computational cost of the RPA-OEP method is about one order of magnitude larger. The initial intent of developing the method was however not to obtain more accurate band energies and gaps but to carry out self-consistent RPA total energy calculations. However, having seen that with the current implementation in GPAW, it was not possible to obtain converged RPA-OEP gaps for the systems with the larger unit cells in Table 5.2 with the computational resources available in a proper amount of time, the initial idea had to be forfeited. Using the resulting RPA-OEP wave functions for self-consistent total-energy calculations of TMOs and more complicated systems was simply deemed impracticable for us. The RPA-OEP potential does differ from the LDA potential in a way leading to a more localized electronic density, due to the fact that the self-interaction errors are reduced. How large the effect is on resulting total energies remain an open question, and given the recent upgrade of our supercomputer, it could be worth revisiting. Another approach to obtaining quasiparticle wave functions would

be to carry out fully self-consistent GW calculations. They could then be used as input for RPA total energies, just as well as the quasiparticle wave functions from the RPA-OEP equation as they are theoretically equivalent. The scGW method is implemented in a handful of codes now, e.g. BerkeleyGW[108], ABINIT[109], West[94] and more, which have attempted to reduce the computational cost by various techniques. With the vertex corrections available in the GW framework as presented in Section 4.3, it would be straight forward to go beyond RPA and obtain $v_{xc}^{\mathrm{rALDA}}(\mathbf{r})$ and from that rALDA wave functions or densities. From the discussion in Section 4.3, one would expect $v_{xc}^{\mathrm{rALDA}}(\mathbf{r})$ to be more positive in both the bonding and interstitial regions in order to raise the energy levels of both the valence and conductions bands, compared with those from the RPA-OEP.

In [104] one positive aspect of the RPA-OEP is argued. They compare their resulting QP band gaps (including the derivative discontinuity) with band gaps of a $\mathrm{GW}^{\mathrm{TC\text{-}TC}}$@HSE method which is a GW QP calculation combining full self-consistency with a vertex included only in the screening (very much similar to the GWP method presented earlier in this work). The comment is that "since two quite different theoretical approaches yield consistently large gaps, one might ask whether the present values are in fact not accurate", with the error being attributed to the lack of electron-phonon interactions in the self-energy. The RPA-OEP method is cheaper than the $\mathrm{GW}^{\mathrm{TC\text{-}TC}}$@HSE as that requires solving the Bethe-Salpeter equation which scales with system size to the power of five, an N^5 scaling. We have however shown in Section 4.3 that only introducing a vertex in the screening is physically not justified and the increased accuracy of the $\mathrm{GW}^{\mathrm{TC\text{-}TC}}$@HSE method is attributed to a fortuitous error cancellation between self-consistency opening the gap and vertex in the screening closing the gap. The agreement between KS gaps from the RPA-OEP method and that of the $\mathrm{GW}^{\mathrm{TC\text{-}TC}}$@HSE method can therefore only be accidental.

CHAPTER 6

Conclusion

In this thesis, improvements and applications of beyond-DFT methods to electronic structure calculations of both bulk and 2D materials were presented. In particular the RPA method was used for total energy calculations in surface science and the GW formalism used to obtain accurate excitations with emphasis on absolute band positions.

The work on the GW method was related to improving descriptions of excitations in both bulk and 2D materials by including vertex corrections using a TDDFT xc kernel. This resulted in the GWΓ method which in particular showed improved IPs and EAs for a range of bulk and 2D semiconductors without increasing the computational cost of the already computationally expensive GW method. The IPs and EAs were shifted upwards by around 0.5 eV. Additionally, slightly larger band gaps were observed for most materials with the GWΓ method, further improving agreement with experimental values. Additionally, the GWΓ method posed a solution to the long-standing problem of the starting-point dependence in GW and improved the numerical convergence behaviour. The GWΓ method was compared with two different approaches of incorporating vertex corrections into the GW method, both however lacking physical justification, and showing a worse agreement with experimental values.

To enable accurate and efficient GW calculations of 2D materials, a correction to the Brillouin zone sampling in the calculation of the self-energy was derived from knowledge of exact limits of the screened Coulomb potential. This reduced the computational time by an order of magnitude without compromising accuracy, thereby enabling large scale calculations. One such large-scale project is the construction of a database of 2D materials. This has already been initiated by the CAMD group at DTU. Here the GWΓ method, enabled by the speed-up of the Brillouin zone integration, will provide accurate band structures enabling users to judge the applicability of a given 2D material in their considered applications.

Some of the GW calculations were employing the new norm-conserving PAW setups that were constructed by using a genetic algorithm. The genetic algorithm replaces the very time consuming hand-tuning of parameters previously necessary to obtain both efficient and accurate PAW setups. Norm-conserving PAW setups is a new addition to the GPAW code, and their importance in GW calculations, especially for accurate prediction of absolute band positions was documented. To which degree the setups of the different elements lack norm-conservation was illustrated and future users of the GW implementation in the GPAW code should now be aware of possible implications that it might have on their calculations. This work enables construction of accurate norm-conserving setups needed for GW calculations.

Besides excitations and the GW method, this work also addressed the lack of accurate data for adsorption energies, important to catalysis, for example. First, the advanced RPA method was applied to calculations of adsorption energies for a range of reactions, where experimental data was available for comparison. Despite the large spread in experimentally reported values, a good agreement was established with the RPA method, deviating with 0.2 eV from the average of the experimental values. The RPA and rALDA methods were subsequently used to construct a database consisting of 200 highly accurate adsorption energies which in turn can be used to guide future development of new xc functionals for DFT, especially useful for surface science. For constructing the family of BEEF functionals, a fitting procedure to a set of trustworthy experimental databases was carried out. Due to the lack of experimental adsorption energy databases, the theoretical one constructed in this work could be added to the fitting procedure, thereby hopefully improving the resulting functional in this regard.
A different approach could be to collect various theoretical databases of highly accurate RPA calculations from the literature. Since RPA is known to perform well across a wide range of different properties, one could attempt to construct the ”RPA-mimicking” DFT functional through a fitting procedure, combining the accuracy of the RPA method with the computational speed of DFT.

With the established accuracy of DFT functionals for adsorption energy calculations in mind, they were applied in the search for catalysts to be used in electrochemical methanol production from ethane. Two different types of surfaces were investigated for this reaction; the (110) surface of rutile transition metal oxides and a fairly new class of two-dimensional materials called

MXenes. Promising candidates were found within the MXenes. Given the vast amount of materials within the class of MXenes and the shown tunability of the adsorption energy by varying the number of layers, a future large-scale study of methanol production is warrented.

CHAPTER 7

Papers

Paper I

A Benchmark database of surface and adsorption energies from many-body perturbation theory

Per S. Schmidt, Thomas Bligaard and Kristian S. Thygesen

In preparation.

A benchmark database of surface and adsorption energies from many-body perturbation theory

Per S. Schmidt, Thomas Bligaard, and Kristian S. Thygesen
Center for Atomic-scale Materials Design (CAMD),
Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark
(Dated: August 31, 2017)

We present an extensive set of surface and adsorption energies calculated using state of the art many-body perturbation theory. In the first part of the paper we consider ten surface reactions in the low coverage regime where experimental data is available. Here the random phase approximation (RPA) is found to yield high accuracy for both adsorption and surface energies. In contrast all the considered density functionals all fail to describe both quantities accurately. This establishes the RPA as a universally accurate method for surface science. In the second part, we use the RPA and the renormalised adiabatic LDA (rALDA) to construct a database of 200 high quality adsorption energies for reactions involving OH, CH, NO, CO, N_2, N, O and H over a wide range of 3d, 4d and 5d transition metals. Due to the significant computational demand, these results are obtained in the high coverage regime where adsorbate-adsorbate interactions can be significant. While RPA and rALDA describe the adsorbate-metal bond similarly, they deviate for the more local adsorbate-adsorbate interactions. The rALDA results are compared to a range of standard density functional theory methods. The deviations are found to be highly functional, surface and reaction dependent. Our work establishes the RPA and rALDA methods as universally accurate full ab-initio methods for surface science where accurate experimental data is scarce. The database is available and should be useful for development of cheaper exchange correlation functionals for surface science

I. INTRODUCTION

The application of density functional theory (DFT) to problems in surface science is ever increasing. In particular, the ability to predict stability and reactivity of transition metal surfaces is an important and fundamental problem in many areas, not least heterogeneous catalysis. Immense efforts have gone into the development of better exchange-correlation (xc)-functionals and today hundreds of different types exist, with the most popular types being the generalized gradient approximations (GGAs), the meta GGAs, (screened) hybrids, GGA+U, and the non-local van der Waals density functionals. With a few exceptions, they all contain parameters that have been optimized for a particular type of problem or class of material. Moreover they rely on fortuitous and poorly understood error cancellation effects. This limits the generality and predictive power of the standard xc-functionals whose performance can be highly system dependent.

Recently, the random phase approximation (RPA) has been advanced as a total energy method that goes beyond standard DFT[1,2]. Within the RPA, the correlation energy is obtained from the linear density response function while exchange is treated exactly. It is computationally much more demanding than conventional DFT (including orbital dependent functionals), but significantly cheaper than wave function-based quantum chemistry methods. The RPA is presently considered the gold standard for solid state systems due to its ab-initio nature, good description of static correlation and excellent account of long range dispersive forces[3,4]. Compared to standard DFT, the RPA reduces self-interaction errors due to its exact treatment of exchange, it incorporates dynamical screening and accounts for long-range correlations such as van der Waals interactions through its non-locality.

In general, the RPA is most accurate for energy differences between isoelectronic systems, i.e. systems with similar electronic structure. Thus structural parameters are generally accurately described as is bonds of dispersive or mixed dispersive-covalent nature. In contrast, strong covalent bond energies are typically underestimated by RPA, implying that atomisation energies of covalently bonded crystals and molecules are systematically underestimated. This deficiency is related to the relatively poor description of the short range correlation hole by the RPA[5,6]. The introduction of an xc-kernel, such as the simple renormalised adiabatic local density approximation (rALDA), in the density response function, greatly improves the short range correlation hole and leads to a significant reduction of the error on covalent bond energies[7–9].

While most previous RPA total energy studies have focused on isolated molecules and bulk solids, there have been some RPA reports on surface adsorption problems[10–15]. Graphene adsorption on metal surfaces is considered notoriously difficult due to the mixed dispersive-covalent nature of the graphene-metal bond. Nevertheless, the predicted RPA binding distances are in excellent agreement with available experimental data[12].The RPA has successfully resolved the "CO adsorption puzzle": in contrast to most DFT functionals RPA predicts the correct binding site of CO on Pt(111) and Cu(111). For the case of CO on Pt(111) and Pd(111), Schimka et al.[10] demonstrated how semi-local xc-functionals underestimate the surface energy relative to experiments and at the same time overestimate the adsorption energy. By modifying the xc-functional, either the predicted adsorption energies or surface energies can

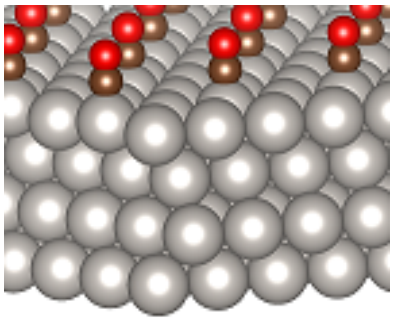


FIG. 1: CO adsorbed in the fcc site at a coverage of 1/4.

be improved but always at the expense of the other[10]. In contrast, the RPA improves the description of both properties simultaneously.

In this work we present an extensive and systematic study of surface adsorption based on the RPA and rALDA methods with comparisons to experiments and selected xc-functionals. The paper consists of two parts. In the first part, we apply the RPA to a range of adsorption reactions at 1/4 coverage where experimental results are available. On average, the RPA predicts adsorption energies that deviate by 0.2 eV from experiments. The same average accuracy is achieved by the RPBE and BEEF-vdW xc-functionals. While the two xc-functionals perform rather similarly, there are significant and non-systematic deviations between the xc-functionals and RPA for the individual reactions indicating the different nature of the RPA compared to the semi-local and vdW density functionals. The good performance of the RPBE and BEEF-vdW for adsorption energies does not carry over to surface energies, which are hugely underestimated. On the other hand RPA remains accurate also for surface energies.

Having established the reliability of the RPA for surface reactions, the second part of the paper is devoted to the development of a comprehensive reference database of adsorption bond energies to transition metal surfaces. The motivation for this endeavour is manifold. First, the database will make it easier for code developers to compare and benchmark their results, which is important in order to enhance the reproducibility of RPA calculations. Important progress along these lines have already started for standard DFT calculations[16]. The much higher complexity of RPA compared to DFT calculations makes it even more important to facilitate such developments for this case. Secondly, the development of better xc-functionals relies crucially on access to large, well defined and consistent datasets. Using experimental data is not ideal because they are influenced by factors not considered in the calculations and thus there is a risk of obtaining the right result for the wrong reason. For methods targeting surface science problems, the situation is even worse because of the scarcity of accurate experimental data for adsorption and surface energies not to mention transition state energies. It is therefore critical to develop theoretical reference datasets for unique reactions and surface structures calculated with the most accurate computational methods available. The concept of theoretical reference databases is widely used in quantum chemistry, but is presently lacking in materials and surface science.

The obvious challenge related to the establishment of RPA (or beyond-RPA) reference datasets for surface science is the large computational cost of such calculations. We overcome this problem by focussing on the high coverage limit where the small size unit cells renders the problem tractable while still permitting an assessment of the critical metal-adsorbate and (adsorbate-adsorbate) interactions. Specifically, we calculate a total of 200 different full coverage adsorption reactions involving OH, CH, NO, CO, N_2, N, O and H adsorbed on a wide range of 3d, 4d and 5d transition metals. We compare the RPA and rALDA results to the PBE[17], RPBE[18] and BEEF-vdW[19] xc-functionals implemented in the GPAW electronic structure code.

PART I: EXPERIMENTALLY RELEVANT REACTIONS

In the first part of this work we use the RPA method to calculate adsorption and surface energies of systems where experimental data is available for comparison[20]. The surface is represented by four layers and the experimentally relevant coverage and adsorption site is used. The reactions considered are CO, NO, O and H on (111)-surfaces of Pt, Rh, Ir, Cu, Pd and Ni. The structures are set up in their PBE-relaxed structures and the RPA calculations are carried out on top of PBE input wave functions and energies. See the Method Section for more details.

The adsorption energy is defined with reference to the corresponding molecule in its gas phase,

$$E_{\mathrm{ads}} = E_{\mathrm{adsorbate@slab}} - (E_{\mathrm{slab}} + E_{\mathrm{adsorbate(g)}})$$

while the surface energy is defined as

$$E_{\mathrm{surf}} = \frac{1}{2}\left(E_{\mathrm{slab}} - N_{\mathrm{layers}} E_{\mathrm{bulk}}\right)$$

A. Adsorption energies

Adsorption energies from PBE, RPBE, BEEF-vdW and RPA and are shown in Table I along with selected experimental values. The experimental reference data are mainly from equilibrium adsorption studies, temperature programmed desorption (TPD) and single crystal adsorption calorimetry (SCAC) (see Ref. [20] and references

		site	PBE	RPBE	BEEF-vdW	RPA	Exp.
CO + Pt(111) →	CO/Pt(111)	top	-1.68	-1.29	-1.20	-1.36	-1.20, -1.22, -1.26, -1.28
CO + Rh(111) →	CO/Rh(111)	top	-1.92	-1.56	-1.50	-1.40	-1.29, -1.33, -1.60
CO + Ir(111) →	CO/Ir(111)	top	-2.11	-1.74	-1.72	-1.48	-1.61
CO + Cu(111) →	CO/Cu(111)	top	-0.74	-0.41	-0.45	-0.25	-0.51, -0.57
CO + Pd(111) →	CO/Pd(111)	fcc	-1.91	-1.48	-1.49	-1.58	-1.34, -1.38, -1.41, -1.43, -1.43, -1.46, -1.47, -1.57
NO + Pd(111) →	NO/Pd(111)	fcc	-2.17	-1.75	-1.75	-1.99	-1.81
NO + Pt(111) →	NO/Pt(111)	fcc	-1.88	-1.45	-1.43	-1.33	-1.16
$\frac{1}{2}O_2$ + Ni(111) →	O/Ni(111)	fcc	-2.16	-1.81	-1.95	-2.50	-2.49
$\frac{1}{2}H_2$ + Ni(111) →	H/Ni(111)	fcc	-0.53	-0.34	-0.27	-0.36	-0.45, -0.48
$\frac{1}{2}H_2$ + Pt(111) →	H/Pt(111)	top	-0.46	-0.28	-0.14	-0.51	-0.36, -0.37
MAE			0.44	0.23	0.19	0.21	-
MSE			-0.38	-0.04	-0.01	-0.10	-

TABLE I

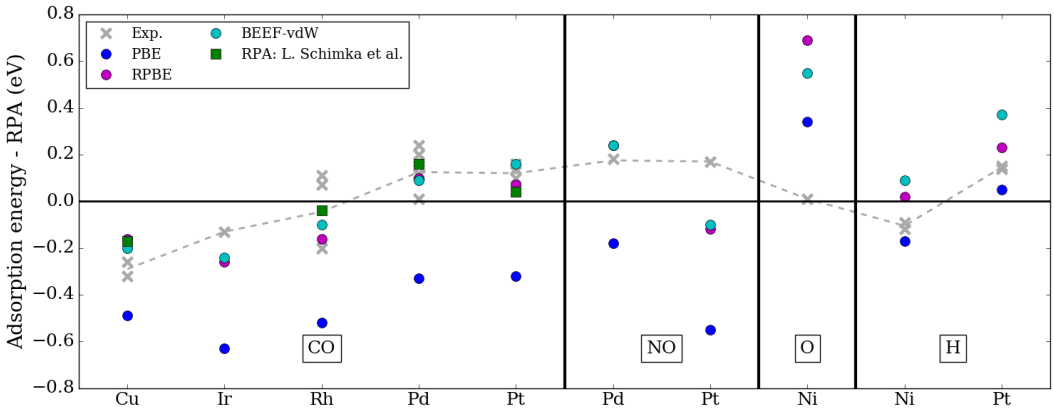


FIG. 2: Deviations in adsorption energies from the RPA results of this work. The green crosses are RPA results from[10].

therein). When comparing to theoretical adsorption energies, it should be kept in mind that, even when carefully executed, such experiments are always subject to inherent uncertainties stemming from variations in the surface crystal structure, presence of surface defects or impurities, different binding sites, side reactions, etc. Consequently, experimental adsorption energies can vary by up to 0.3 eV for the same reaction although for most reactions is around 0.1 eV. The mean absolute error (MAE) and mean signed error (MSE) of the theoretical results have been calculated relative to the average experimental adsorption energy reported for each reaction.

Fig.2 shows the deviations of the calculated and experimental adsorption energies from the RPA values of this work. The RPA, RPBE, and BEEF-vdW all have a MAE of around 0.20 eV which is significantly lower than the 0.44 eV obtained with PBE. The systematic underestimation of adsorption energies by the PBE is a well known problem. The case of O on Ni(111) deviates from the general trends which is related to the rather poor description of the O_2 reference by the DFT functionals. It should be noted that the BEEF-vdW has been fitted to a dataset containing the first 8 adsorption reactions considered.

The performance of the RPA is actually quite remarkable in view of its pure ab-initio nature. With a MSE of -0.10 eV, it appears that the RPA has a weak tendency to overbind the adsorbates. However, given the small size of the dataset and the inherent uncertainties in experimental data one should be careful to draw too strict conclusions. The agreement with previously published RPA values (for CO on Cu, Rh, Pd, Pt) is acceptable with a MAE of 0.10 eV and MSE of 0.002 eV.

These deviations could stem from differences in the applied geometries (although PBE relaxed structures were used in both studies) and the fact that the RPA results

from Ref. 10 were obtained by extrapolating RPA energies for a $\sqrt{3} \times \sqrt{3}$ unit cell to a 2×2 cell using PBE data.

B. Surface energies

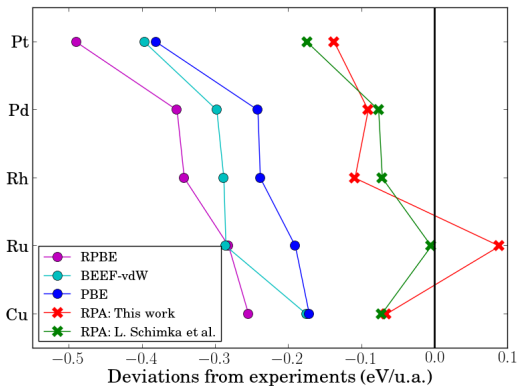


FIG. 3: Surface energies of five different fcc(111) surfaces calculated with four different methods and compared to experimental values and RPA values from [10].

Surface energies of some selected (111) surfaces were calculated with RPA, rALDA and different DFT xc-functionals. Fig. 3 shows the surface energies obtained in the present work together with the RPA results of Ref. [10], plotted relative to the experimental values. Our RPA results show excellent agreement with the previous RPA values and confirm that RPA predicts surface energies in much better agreement with experiments than all the considered DFT functionals. Inclusion of the rALDA kernel does not have a large effect on the surface energies which are very similar to the RPA values (not shown).

As previously observed, it is striking that the xc-functionals which perform better for adsorption energies (RPBE and BEEF-vdW) perform worse for surface energies and vice versa. This circumstance highlights the limitations of presently employed functionals and underlines the advantage of more advanced methods, like the RPA and the rALDA. The latter exhibit a greater degree of universality in the sense that they offer high (though not perfect) accuracy across a broad range of systems and bonding types.

C. Coverage effects

As mentioned in the introduction, the computational cost of the RPA and rALDA makes large scale application of these methods a daunting task. As a consequence, for benchmarking purposes it is desirable to explore minimalistic models of adsorption systems which still capture the essential physical mechanisms governing surface reactions, i.e. the formation and breaking of bonds between an extended metal surface and the adsorbate. In Fig. 4 we show the coverage dependence of the adsorption energy for the case of CO on top of a Pt(111) surface. Cleary, there is a significant dependence of the binding energy on the coverage due to repulsive adsorbate-adsorbate interactions at higher coverages. However, the relative ordering of the adsorption energy obtained with different methods is essentially unchanged. This shows that adsorption in the high coverage limit involves much the same physics as the low coverage regime. Having made this point, it should be stressed that the high coverage regime is of interest in itself. In indeed, high coverage configurations become relevant at high gas pressure conditions. Moreover, adsorbate-adsorbate interactions (which contribute to the adsorption energy at high coverage) are generally important to incorporate for a correct description of reaction kinetics. Finally, high coverage configurations do, at least to some extent, resemble the stretched bond configurations of dissociative transition states. Thus one could expect that the more accurate description of high coverage configurations entails more accurate description of transition states and barrier heights. In this context, it is interesting to note from Fig. 4 that the magnitude of the repulsive adsorbate-adsorbate interactions is largest with the pure GGAs (PBE and RPBE), smaller with the van der Waals density functional (BEEF-vdW), and smallest with the RPA. We believe this is due to the stabilising effect of attractive van der Waals interactions between the adsorbates at higher coverage.

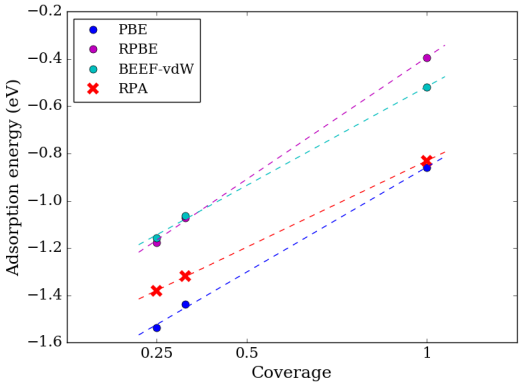


FIG. 4: Adsorption energies of CO on Pt(111) with a coverage of 1/4 to 1 calculated with various methods. The binding energy is decreased at higher coverage due to repulsive adsorbate-adsorbate interactions, however, the relative ordering of the adsorption energy obtained with different methods stays more or less unchanged. The magnitude of the adsorbate-adsorbate interactions is smaller with the RPA than the DFT functionals due to the stabilizing effect of the attractive van der Waals interactions between the molecules at higher coverage.

PART II: DATABASE OF ADSORPTION ENERGIES

In this section we present 200 adsorption energy calculations of full coverage reactions in order to compare the performance of different computational methods. The adsorption is on the top site of the fcc(111) surface represented by three layers using the bulk PBE lattice constant. The position of the top surface layer and the adsorbate has been relaxed with PBE.

The adsorption energies have been calculated with PBE, RPBE, BEEF-vdW and their relative performance compared to that of the rALDA.

The adsorption energies were calculated for the following reactions on transition metal surfaces with the reference molecule in their gas phase:

$$
\begin{aligned}
&(1) && H_2O + \text{slab} \rightarrow OH/\text{slab} + \tfrac{1}{2}\, H_2\\
&(2) && CH_4 + \text{slab} \rightarrow CH/\text{slab} + \tfrac{3}{2}\, H_2\\
&(3) && NO + \text{slab} \rightarrow NO/\text{slab}\\
&(4) && CO + \text{slab} \rightarrow CO/\text{slab}\\
&(5) && N_2 + \text{slab} \rightarrow N_2/\text{slab}\\
&(6) && \tfrac{1}{2}N_2 + \text{slab} \rightarrow N/\text{slab}\\
&(7) && \tfrac{1}{2}O_2 + \text{slab} \rightarrow O/\text{slab}\\
&(8) && \tfrac{1}{2}H_2 + \text{slab} \rightarrow H/\text{slab}
\end{aligned}
$$

An example of the full coverage adsorption is shown in Fig. 5. The adsorbate-adsorbate distance is around 2.5 Å and adsorbate-adsorbate interactions will therefore certainly affect the adsorption energies. The adsorption energy essentially consists of two contributions: Firstly, bringing the individual molecules in close proximity forming a molecular "layer" and secondly adsorbing the full-coverage layer on the surface. These two different contributions are calculated with RPA and rALDA and shown in Fig. 6 for CO adsorbed on Mn(111). The two methods agree well for the adsorption of the molecular layer with differences below 0.1 eV across all adsorbates. When dissociating the layer however, RPA and rALDA are shown to differ quite significantly (up to 0.42 eV for NO). The differences between RPA and rALDA for all adsorbates on a few different surfaces are shown in Fig. 7. The difference is seen to be surface independent.

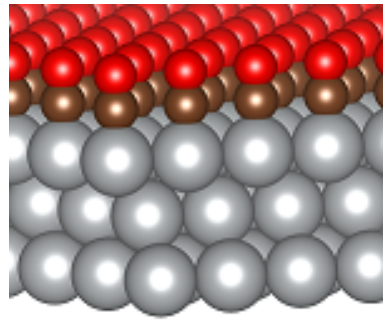


FIG. 5: CO adsorbed in the top site at full coverage.

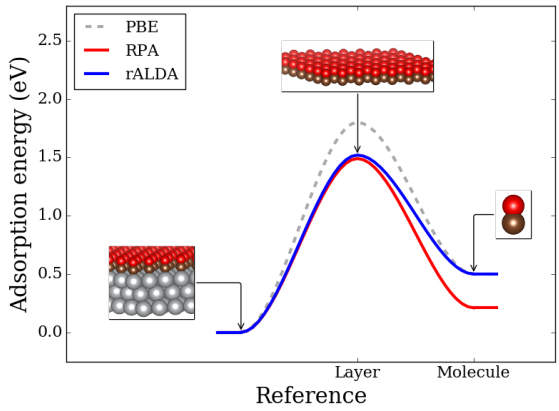


FIG. 6: RPA and rALDA calculations of CO adsorbed on Mn(111) with respect to the molecular layer and the isolated CO molecule.

D. Results

1. *Adsorption energies*

The differences between PBE, RPBE, BEEF-vdW and RPA are shown in Fig. 8. Blue (red) represents overbinding (underbinding). PBE performs quite well with an average deviation across all reactions of 0.16 eV. The underbinding is worse on the later transition metals and the adsorption of CH and O are described the worst. The sign of the differences are seen to be reaction dependent.

BEEF-vdW predicts weaker binding for almost all reactions. The same was observed for 1/4 coverage in the first part of this work. The average deviation is 0.22 eV across all reactions. The descriptions of CO and O adsorption are the worst. RPBE predicts an even weaker binding for almost all reactions. The deviations vary from 0.09 eV for the adsorption of H to 0.53 in the case of the adsorption of N_2. In average the difference is 0.35 eV.

The adsorption energy is shown vs surface energy for two different reactions on four different surfaces in Fig. 9 calculated with five different methods. The DFT results fall roughly on a straight line, as shown previously[10], supporting the statement that by tuning the xc functional, we change both adsorption and surface energies simultaneously. rALDA is seen to deviate from this line significantly and supposedly corrects both energies. The magnitude of this deviation is seen to be surface and reaction dependent. In some cases the BEEF-vdW results deviate a bit from this line as well, suggesting that it is

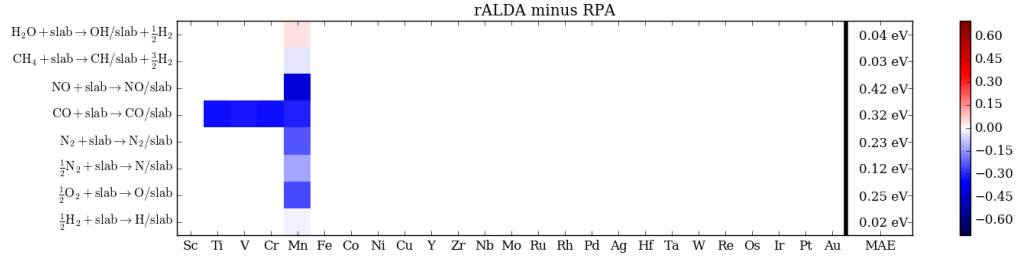


FIG. 7: Difference in adsorption energies calculated with RPA and rALDA with the isolated molecule as reference.

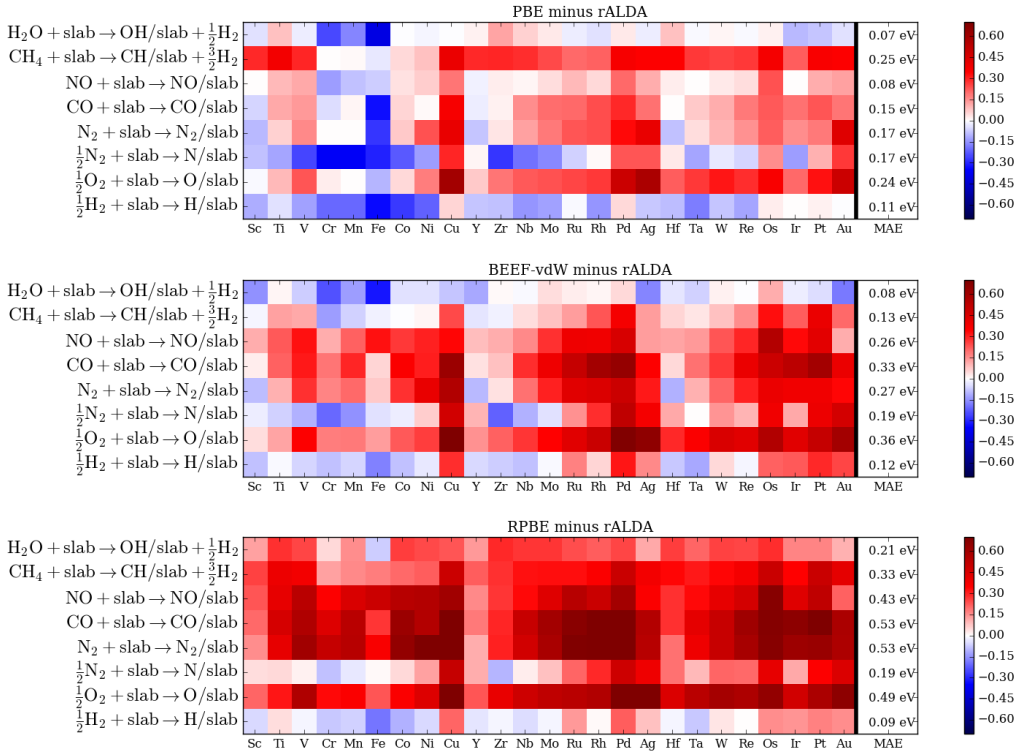


FIG. 8: Difference in adsorption energies calculated with PBE, RPBE and BEEF vs rALDA with respect to the isolated molecule.

indeed the incorporation of van der Waals forces that is important.

The surface energy is shown versus the adsorption energy for the 8 different reactions on Fig. 10. The type of calculation is represented by a different color and the length and direction of the arrows show the difference between PBE (red), BEEF-vdW (green) and rALDA (blue). The direction of the arrow varies between the different reactions. In average, the surface is predicted to be too stable by the same amount with both PBE and BEEF-vdW.

CONCLUSION

In conclusion we have established the RPA and rALDA methods as universally accurate full ab-initio methods for surface science. A database of 200 high quality, full cov-

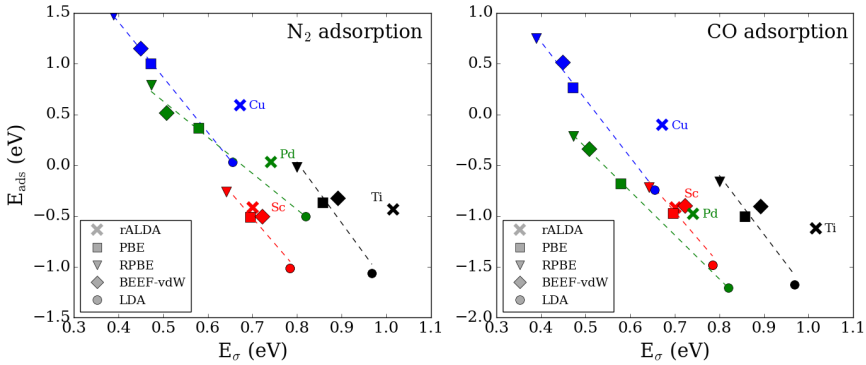


FIG. 9: Comparing adsorption and surface energies calculated with rALDA (cross), PBE (square), RPBE (triangle), BEEF (diamond) and LDA (circle) for the adsorption of O and NO on four different surfaces.

erage, adsorption energies was constructed and compared with existing DFT functionals; the PBE, RPBE and BEEF-vdW. The RPA and rALDA methods were seen to differ quite significantly in the low-coverage regime which mimicks the observed differences in atomization energies as well.

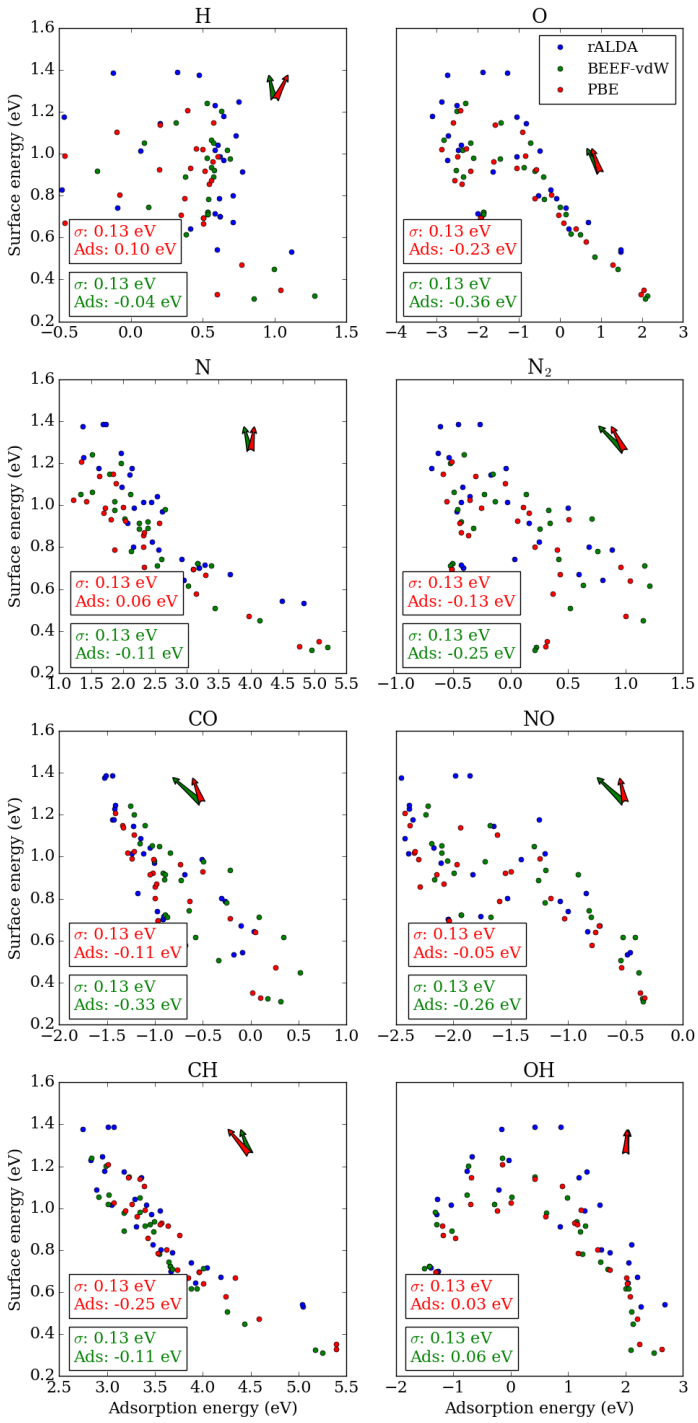


FIG. 10: The surface energy is plotted against the adsorption energy calculated with rALDA (red), BEEF-vdW (green) and PBE (blue). Each dot represents a surface. The size and direction of the green (blue) arrow shows the difference between RPA and BEEF-vdW (PBE). The numbers in the boxes correspond to the length of the arrow in the x and y direction.

II. METHOD SECTION

A. Computational implementation: The random phase approximation

We employ the implementation of RPA in the electronic structure code GPAW[21], as described in greater detail in **?** . In short, within RPA the xc energy contribution to the total ground state energy is split into an exact-exchange term[?] plus RPA correlation. The exact exchange term can be written using plane waves as:

$$E_x = -\frac{1}{N_{\mathbf{q}} N_{\mathbf{k}}} \sum_{n,n'} \sum_{\mathbf{k},\mathbf{q}}^{1.\mathrm{BZ}} f_{n\mathbf{k}} \theta(\epsilon_{n\mathbf{k}} - \epsilon_{n'\mathbf{k}+\mathbf{q}}) \quad (1)$$

$$\sum_{\mathbf{G}} v_{\mathbf{G}}(\mathbf{q}) |\langle \psi_{n\mathbf{k}} | e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} | \psi_{n'\mathbf{k}+\mathbf{q}} \rangle|^2 \quad (2)$$

where the number of plane waves is determined by a cutoff energy, $|\mathbf{q}+\mathbf{G}|^2/2 < E_{\mathrm{cut}}$. The correlation energy contribution to the total energy is calculated from the non-interacting response function, $\chi^0(i\omega)$, by:

$$E_c^{\mathrm{RPA}} = \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\Big[\ln(1 - \chi^0(i\omega)v) + \chi^0(i\omega)v \Big] \quad (3)$$

where v is the Coulomb interaction and Tr is the trace. The response function and Coulomb interaction are evaluated in a plane wave basis and for periodic systems the trace involves a summation over **q**-points, which are determined from the Brillouin zone sampling. The frequency integration is carried out using a Gaussian quadrature[?] . In a plane wave basis, the non-interacting response function is given as

$$\chi^0_{\mathbf{G},\mathbf{G}'}(\mathbf{q}, i\omega) = \frac{1}{\Omega} \sum_{\mathbf{k}}^{\mathrm{BZ}} \sum_{n,n'} \frac{f_{n\mathbf{k}} - f_{n'\mathbf{k}+\mathbf{q}}}{i\omega + \epsilon_{n\mathbf{k}} - \epsilon_{n'\mathbf{k}+\mathbf{q}}} \times \langle \psi_{n\mathbf{k}} | e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} | \psi_{n'\mathbf{k}+\mathbf{q}} \rangle \langle \psi_{n'\mathbf{k}+\mathbf{q}} | e^{i(\mathbf{q}+\mathbf{G}')\cdot\mathbf{r}'} | \psi_{n\mathbf{k}} \rangle \quad (4)$$

where Ω is the volume of the unit cell and $f_{n\mathbf{k}}$ occupation numbers. The sum over states is in principle infinite but in practice truncated at a finite number determined by the cutoff energy and the resulting total energy is then extrapolated using the usual $1/E_{cut}^{3/2}$ scheme

$$E_c = E_c^\infty + \frac{A}{E_{cut}^{3/2}}$$

B. Computational details

1. Experimental comparison

The surfaces are represented by four layers with the bottom two fixed by the bulk PBE lattice constant found in the supplementary material of [20], and the top two layers relaxed together with the adsorbate. 10Å of vacuum was used between neighbouring slabs and the reference is the isolated spin-polarized molecule in a 6Å × 6Å × 6Å box. The RPA calculations were carried out using $8 \times 8 \times 1$ **k**-points and an identical q-point grid for Brillouin zone integration. 16 frequency points were used for the frequency integration and the extrapolation to infinite cutoff was done from calculations at 200, 250 and 300 eV. The EXX and RPA energies were evaluated on top of PBE eigenvalues and orbitals. The calculations are carried out using the plane wave version of the Grid-based Projector-Augmented Wave method (GPAW) code[21] and the Atomic Simulation Environment (ASE)[?] with the associated 0.9.11271 atomic setups.

2. Adsorption database

The surfaces were modeled using three layers with the bottom two layers fixed at the fcc PBE lattice constants[?] and the position of the top layer relaxed. See the supplementary material for information on how the adsorption energy is affected by the slab thickness. The position of the adsorbate was relaxed keeping all three surface layers fixed at the position found previously. All relaxations were carried out with the BFGS algorithm using the PBE approximation to the xc-functional with a force convergence criteria of 0.05 eV/Å. The electron temperature was 0.01 eV and spin-polarized calculations were performed for calculations involving Fe, Ni or Co. 5 **Å** of vacuum was added to either side of the adsorbate to avoid artificial interactions between neighboring layers following convergence tests at both the DFT and RPA level. The adsorption energies are relative to the molecule in its gas phase and the calculations for the isolated molecules were carried out in a $7 \times 7 \times 7$ Å^3 box fully relaxing the geometry with the PBE functional.

The RPA calculations were carefully converged with respect to plane wave basis using the following extrapolation scheme: In Fig. 11 the black dots are from a calculation with $6 \times 6 \times 1$ k-points (not enough to achieve convergence) but high cutoff energies (300, 400, 500 eV). The green circle is a calculation at a much denser k-point sampling of $12 \times 12 \times 1$ (converged). From these four circles, the two green crosses are predicted which allow for an extrapolation to infinite cutoff energy. The red dots represent actual calculations with both a dense **k**-point grid and high cutoff energies to test the extrapolation scheme. The error introduced by the extrapolation scheme for this particular system is seen to be 0.013 eV. The **k**-point grid of $12 \times 12 \times 1$ ensures that the exchange + correlation energy is converged to within 0.02 eV with respect to the **k**-point density.

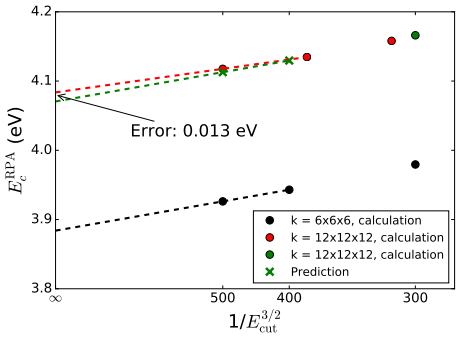


FIG. 11: Correlation energy contribution to the total adsorption energy of O adsorbed on Mn. The black dots are from a calculation with $6 \times 6 \times 1$ k-points (not converged) but high cutoff energies (300, 400, 500 eV). The green circle is a calculation at a much denser k-point sampling of $12 \times 12 \times 1$ (converged). From these four circles, the two green crosses are predicted which allow for an extrapolation to infinite cutoff energy. The red dots represent actual calculations with both a dense **k**-point grid and high cutoff energies to test the extrapolation scheme. The error introduced by the extrapolation scheme for this particular system is seen to be 0.013 eV.

ACKNOWLEDGMENTS

We acknowledge support from the Danish Council for Independent Research's Sapere Aude Program, Grant No. 11-1051390. The Center for Nanostructured Graphene is sponsored by the Danish National Research Foundation, Project DNRF58.

[1] F. Furche, Phys. Rev. B **64**, 195120 (2001).
[2] M. Fuchs and X. Gonze, Phys. Rev. B **65**, 235109 (2002).
[3] X. Ren, P. Rinke, C. Joas, and M. Scheffler, Journal of Materials Science **47**, 7447 (2012).
[4] A. Marini, P. García-González, and A. Rubio, Phys. Rev. Lett. **96**, 136404 (2006).
[5] L. Hedin and S. Lundqvist, Solid State Physics **23**, 1 (1970).
[6] K. S. Singwi, M. P. Tosi, R. H. Land, and A. Sjölander, Phys. Rev. **176**, 589 (1968).
[7] T. Olsen and K. S. Thygesen, Phys. Rev. Lett. **112**, 203001 (2014).
[8] T. Olsen and K. S. Thygesen, Phys. Rev. B **88**, 115131 (2013).
[9] T. Olsen and K. S. Thygesen, Phys. Rev. B **86**, 081103 (2012).
[10] L. Schimka, J. Harl, A. Stroppa, A. Grüneis, M. Marsman, F. Mittendorfer, and G. Kresse, Nature Materials **9**, 741 (2010).
[11] X. Ren, P. Rinke, and M. Scheffler, Phys. Rev. B **80**, 045402 (2009).
[12] T. Olsen, J. Yan, J. J. Mortensen, and K. S. Thygesen, Phys. Rev. Lett. **107**, 156401 (2011).
[13] M. Rohlfing and T. Bredow, Phys. Rev. Lett. **101**, 266106 (2008).
[14] J. Ma, A. Michaelides, D. Alfè, L. Schimka, G. Kresse, and E. Wang, Phys. Rev. B **84**, 033402 (2011).
[15] H.-J. Kim, A. Tkatchenko, J.-H. Cho, and M. Scheffler, Phys. Rev. B **85**, 041403 (2012).
[16] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso, S. de Gironcoli, T. Deutsch, J. K. Dewhurst, I. Di Marco, C. Draxl, M. Dułak, O. Eriksson, J. A. Flores-Livas, K. F. Garrity, L. Genovese, P. Giannozzi, M. Giantomassi, S. Goedecker, X. Gonze, O. Grånäs, E. K. U. Gross, A. Gulans, F. Gygi, D. R. Hamann, P. J. Hasnip, N. A. W. Holzwarth, D. Iuşan, D. B. Jochym, F. Jollet, D. Jones, G. Kresse, K. Koepernik, E. Küçükbenli, Y. O. Kvashnin, I. L. M. Locht, S. Lubeck, M. Marsman, N. Marzari, U. Nitzsche, L. Nordström, T. Ozaki, L. Paulatto, C. J. Pickard, W. Poelmans, M. I. J. Probert, K. Refson, M. Richter, G.-M. Rignanese, S. Saha, M. Scheffler, M. Schlipf, K. Schwarz, S. Sharma, F. Tavazza, P. Thunström, A. Tkatchenko, M. Torrent, D. Vanderbilt, M. J. van Setten, V. Van Speybroeck, J. M. Wills, J. R. Yates, G.-X. Zhang, and S. Cottenier, Science **351** (2016), 10.1126/science.aad3000, http://science.sciencemag.org/content/351/6280/aad3000.full.pdf.
[17] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[18] B. Hammer, L. B. Hansen, and J. K. Nørskov, Phys. Rev. B **59**, 7413 (1999).
[19] J. Wellendorff, K. T. Lundgaard, A. Møgelhøj, V. Petzold, D. D. Landis, J. K. Nørskov, T. Bligaard, and K. W. Jacobsen, Phys. Rev. B **85**, 235149 (2012).
[20] J. Wellendorff, T. L. Silbaugh, D. Garcia-Pintos, J. K.

Nørskov, T. Bligaard, F. Studt, and C. T. Campbell, Surface Science **640**, 36 (2015), reactivity Concepts at Surfaces: Coupling Theory with Experiment.

[21] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, Phys. Rev. B **71**, 035109 (2005).

Paper II

Electrochemical methanol production from methane

Logi Arnarson, Per S. Schmidt, Kristian S. Thygesen, Ifan E. L. Stephens and Jan Rossmeisl

In preparation.

Electrochemical Methanol Production from Methane

Logi Arnarson and Jan Rossmeisl
Nano-Science Center, Department of Chemistry, University of Copenhagen, Universitetsparken 5, Copenhagen, Denmark

Per S. Schmidt and Kristian S. Thygesen
Center for Atomic-scale Materials Design (CAMD), Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

Ifan E. L. Stephens
Department of Materials, Imperial College London, Exhibition Road, London SW7 2AZ, UK
(Dated: August 31, 2017)

In this work, the production of methanol via an electrochemical process is found to be viable by considering a range of different rutile MeO_2(110) surfaces and $AB_2C_2O_2$ MXenes as electrocatalysts. Methane is oxidized to methanol with the overall reaction $CH_4 + H_2O \rightarrow CH_3OH + H_2$ and the reaction pathway is here determined using expensive CI-NEB calculations for a few systems. The resulting activation energy is found to be well described by a simple descriptor, enabling future large scale screening studies. The required high selectivity towards methanol production instead of forming O_2 puts an upper limit on the applied potential, according to the full OER pathway. Some promising candidates were found that showed a sufficiently low reaction energy barrier whilst not breaking this upper limit.

I. INTRODUCTION

Every day, an estimated 5.3 trillion cubic feet of methane (natural gas) are being burned as waste at oil fields around the globe[1]. This sheer volume equates to 20% of the U.S. gas consumption[2] and 3% of the global CO_2 emissions[3]. Every year, an enormous amount of energy (30 EJ[3]) is lost when methane (CH_4, natural gas) is flared in oil fields, releasing CO_2 into the atmosphere. Flaring occurs at oil fields in remote locations, where there is a lack of infrastructure to store and transport the gaseous methane. This huge waste of a precious and finite resource highlights an opportunity for remote-deployable technologies to harness the energy of waste methane and store it in a more portable form. Methane has 25 times the global warming potential of carbon dioxide, which is why it is usually burned instead of released directly[4]. However, it could be far more attractive to oxidise methane into a liquid fuel, such as methanol, which would fit better in to the existing infrastructure. This challenge has captivated researchers in the heterogeneous catalysis community for decades[5,6]. However, the direct routes to produce methanol from methane are inefficient, require high pressures, are unselective or require oxidants such as H_2O_2, which is notoriously challenging to transport to remote locations. More complex, indirect multistage processes can yield higher efficiencies, but require high temperatures and pressures and hence large amounts of infrastructure[7]. In contrast, the electrochemical oxidation of methane to methanol could be far more attractive. A schematic of an electrochemical fuel cell producing methanol is shown on Fig. 1. The only inputs would be methane, oxygen (from air) and water. The overall reaction is

$$CH_4 + H_2O \rightarrow CH_3OH + H_2, \quad \Delta G \cong 120\,\mathrm{kJ/mol}. \tag{1}$$

At the anode, CH_4 would be oxidised to methanol according to:

$$CH_4 + H_2O \rightarrow CH_3OH + 2(H^+ + e^-) \tag{2}$$

whereas at the cathode, O_2 would be reduced to H_2O:

$$\frac{1}{2}O_2 + 2(H^+ + e^-) \rightarrow H_2O \tag{3}$$

Rather than consuming energy, such a device could in principle produce energy to harness in the form of electricity, in addition to the liquid fuel methanol. This is the ultimate solution for tapping into waste methane. The device would be portable with easy scale up, the produced liquid methanol could be easily stored in tanks in areas without sufficient pipelines. Missing is a method to produce methanol at high rates and selectively avoiding the formation of the more thermodynamically favoured product, CO_2, according to:

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O, \tag{4}$$

$$\Delta G \cong -820\,\mathrm{kJ/mol}. \tag{5}$$

As for CO_2 reduction, the key component in controlling the performance of electrochemical devices is the interface between the electrolyte and electrode. Typically, metal electrocatalysts produce CO_2 upon oxidising methane. A number of reports in the literature have been devoted to the oxidation of methane to methanol at low temperatures (i.e. 100°C or below), via (i) an indirect method or (ii) a direct method. In the indirect method,

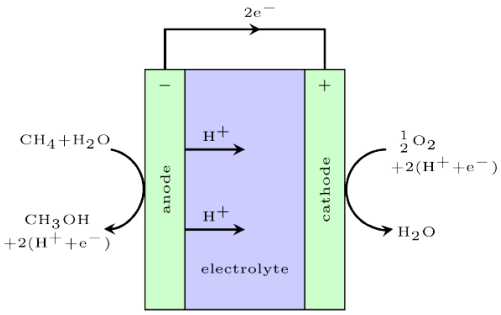


FIG. 1: Schematic figure of a device converting CH_4 and O_2 into CH_3OH.

the same electrode acts as cathode and anode, reducing O_2 and at the same time oxidising methane[8,9]. Earlier studies suggest that this proceeds by the formation of the superoxide anion via an "outer-sphere" solution phase step[8]: $O_2 + e^- \rightarrow O_2^-$. This highly reactive species would abstract a proton from methane, allowing it to be oxidised. However, it would be far more attractive to oxidise methane directly, avoiding its potentially explosive mixture with O_2.

Upon oxidising methane, reactive metals such as Pt produce close to 100% CO_2[10]. On the other hand, oxides seem to be able to produce methanol; for instance V_2O_5 on SnO_x produces 60% methanol at a current density of 4 mA cm^{-2} at 100°C[11]. Methanol production peaked at a cell potential of 0.9 V, i.e. equivalent to an energy input of ∼170 kJ/mol; however, this input is only 25% of the Gibbs free energy released upon the complete oxidation of methanol to CO_2. In order for this to be industrially viable, we would need to increase the current density, improve the selectivity and minimize the overpotential.

Our hypothesis is that the catalyst material needs to bind oxygen weakly or as a radical, to have the driving force needed to oxidize methane. The consequence is that the potential needed to run the reaction: $H_2O + * \rightarrow *O + 2(H^+ + e^-)$ is high, but at the same time below the potential necessary to catalyse the evolution of O_2. Controlling the selectivity towards methanol versus more thermodynamically favoured products, in particular CO_2, would be challenging. Based on the work of Behm and coworkers on similar reactions, we propose that accelerated mass transport of methanol away from the surface could prevent its oxidation[12]. This could be controlled by judicious electrode structuring and mass transport.

In this study we investigate two different surfaces for methanol production. The (110) surface of rutile transition metal oxides, MeO_2 (Me; a transition metal atom), and monolayer MXenes of the $AB_2C_2O_2$ type (A, B: two different transition metal atoms). MXenes is a new class of two-dimensional materials comprising carbides and nitrides of transition metals. MXenes have shown promise for various applications in lithium/sodium ion batteries and as HER catalysts[13]. MXenes come in various types; M_2X, M_3X_2, M_2XA_2, $M_3X_2A_2$, $M_4X_3A_2$ etc., where M refers to the transition metal, X refers to carbon or nitrogen and A refers to the functional group present on the surface (O, OH or F). In this study, we limit ourselves to the metallic $AB_2C_2O_2$ type, which is shown in Fig.2. The rationale to use O as the functional group is the solvothermal synthesis of MXenes under acidic condition where O and OH are present to passivate the surface[14]. Previous studies have found that functionalization with O is thermodynamically preferred[15]. The same study investigated the use of some of the MXenes for HER. They found that the hydrogen adsorption energy varied by up to 0.5 eV depending on the number of metal layers suggesting that catalytic activity can be tuned by controlling layer thickness.

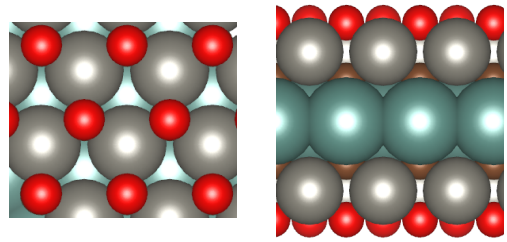


FIG. 2: Top and side view of $YW_2C_2O_2$.

Given the possibility to tune the number of layers and the vast choices of M, X and A, this class of materials is huge. In this study we limit ourselves to seven candidates.

II. RESULTS AND DISCUSSION

The first part of the electrochemical production of methanol from methane is to split water by means of electrochemistry in a water electrolyser to provide an adsorbed oxygen atom. This is also known as the first two elementary reactions in the oxygen evolution reaction (OER). All four reactions of the OER are:

$$* + 2H_2O \rightarrow *OH + H_2O + (H^+ + e^-) \quad (6)$$
$$\rightarrow *O + H_2O + 2(H^+ + e^-) \quad (7)$$
$$\rightarrow *OOH + 3(H^+ + e^-) \quad (8)$$
$$\rightarrow * + O_2 + 4(H^+ + e^-), \quad (9)$$

where $*$ represents an active site on the surface. In Fig. 3 the full OER energetics, according to reactions (5)-(8), for a range of different MeO_2(110) surfaces is shown (Me being a transition metal). It is the stability of the $*O$ intermediate which is of importance for methanol production since this is the oxygen atom that will oxidise methane to methanol. It is clear from Fig. 3 that the

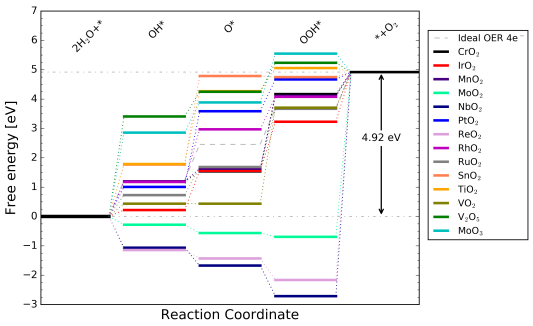


FIG. 3: OER free energy diagram for the MeO_2 class.

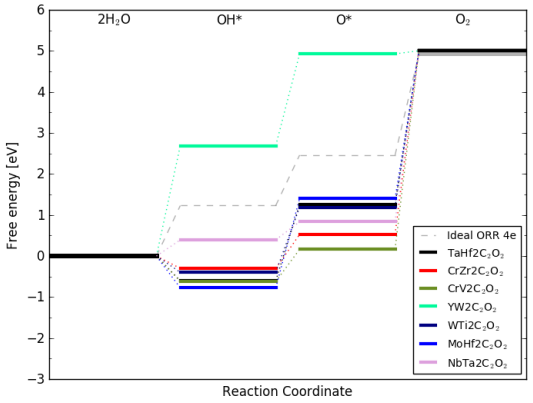


FIG. 4: OER free energy diagram for the MXenes.

different materials provide an oxygen atom with vastly different degree of reactivity. For NbO_2 it is energetically advantageous to form $*O$ by -1.41 eV whereas it costs 3.54 eV on PtO_2. In Fig. 4 the same free energy diagram is shown for the MXenes, omitting the $*OOH$ intermediate. Different reactivities are found from this class of materials as well. With an oxygen atom adsorbed on the surface, the second part of methanol synthesis is having the oxygen react with methane. This is a purely heterogeneous catalytic step. The reaction pathway consists of the following four steps

$$*O + CH_4(g) \rightarrow *O\cdots CH_4 \quad (10)$$
$$\rightarrow *OH\cdots CH_3 \quad (11)$$
$$\rightarrow *CH_3OH \quad (12)$$
$$\rightarrow * + CH_3OH(g) \quad (13)$$

where $*O$ in Eq. (10) is the same site as $*O$ introduced in Eq. (7). The product of the reaction in Eq. (10) is an oxygen atom bonded to the active site and CH_4 in the vicinity to it. Eq. (11) denotes the transition state and Eq. (12) is CH_3OH bonded to the surface. Eq. (13) is the final state, which is the clean surface and gas phase methanol. This reaction pathway was confirmed by carrying out the expensive CI-NEB calculations for RhO_2, see Fig. 5 and a few of the MXenes. In Fig. 5, images are shown along the pathway. The first image in Fig. 5 corresponds to $*O\cdots CH_4$ in Eq. (10) and the last to $*CH_3OH$ in Eq. (12). Image 3 in Fig. 5 is the transition state where the methane has delivered a single H atom to the $*O$ site and CH_3 is in the vicinity. This transition state is in consensus with what has been named the radical pathway in a recent publication from Nørskov et. al.[16]

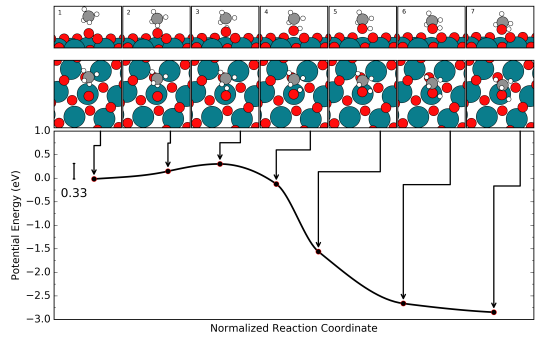


FIG. 5: Reaction pathway calculated using the CI-NEB method on RhO_2 for methane oxidation.

Analysis of the structure of the adsorbate in the transition state shows that the CH_3 molecule exhibits planar geometry with all H-C-H bonding angles ∼120°. Whereas the geometry of the CH_4 molecule is tetrahedral with all H-C-H bonding angles of ∼109.5°. This suggests that the energy required to break the first C-H bond, to activate methane, equals the energy needed to deliver a single H atom to the surface and form CH_3 from CH_4, i.e.:

$$E_{\text{act}}^{\text{estimated}} = \left(E_{CH_3(g)} - E_{CH_4(g)}\right) + \left(E_{*OH} - E_{*O}\right) \quad (14)$$

This corresponds to converting the sp3 hybridisation in CH_4 to sp2 hybridisation in CH_3 and the double bonded oxygen to the metal atom to a single bonded OH group on the surface. In Fig. 6 the activation energy of Eq. (14) is plotted with circles against the descriptor $E_{*OH} - E_{*O}$. The MXenes are plotted in light green. With this model, all materials lie on a straight line. The activation energy calculated with the CI-NEB method is marked with squares in Fig. 6.Marked with triangles is an approximated activation energy using the already found CI-NEB path and unit cell of RhO_2 for all the other materials of the MeO_2 type. The approximated and calculated activation energies are found to be in reasonably good agreement with the estimated values using the very simple expression in Eq. (14). For RhO_2, IrO_2, RuO_2 and CrO_2 the estimate is excellent with a deviation below 0.05 eV from the calculated value. For ReO_2 and NbO_2 the deviation is slightly below 0.20 eV (overestimated for the

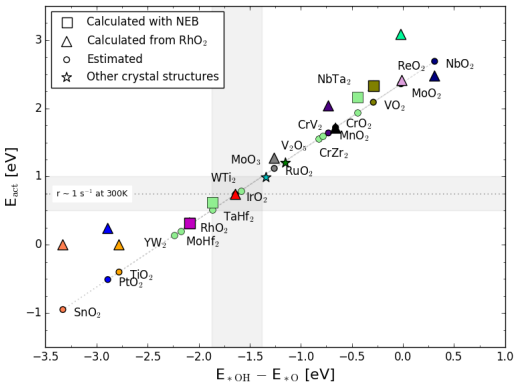


FIG. 6: The activation energy of methane activation plotted as a function of the descriptor $E_{*\mathrm{OH}} - E_{*\mathrm{O}}$. The MXenes are plotted in light green.

former and underestimated for the latter) and for MnO_2, VO_2 and MoO_2 the deviation is above 0.30 eV. For SnO_2, PtO_2 and TiO_2 the estimation for the activation energy is below 0 eV i.e. indicating formation of has no kinetic barrier. The calculated activation energy is found to be very low or ∼0 eV in agreement with that. Fig. 6 can also be interpreted as the potential vs. rate i.e. the applied potential needed to run the reaction vs. rate of methanol production. As higher potential is applied (more negative $E_{*\mathrm{OH}} - E_{*\mathrm{O}}$ values) the lower the energy barrier for methanol formation becomes and vice versa. As a rule of thumb for gas phase chemistry to obtain a rate of ∼1 s^{-1} at 300K the energy barrier should be ∼0.75 eV. A few materials fall at or below this value. To ensure high selectivity towards CH_3OH instead of forming O_2, only a moderate potential should be applied, according to the full OER pathway. This will come at the cost of a higher activation energy. With this in mind, promising candidates could be IrO_2, $MoHf_2C_2O_2$ and $TaHf_2C_2O_2$ since the reaction barriers are close to or below 0.75 eV and the potential necessary to adsorp oxygen on the surface is not sufficient for the oxygen to form O_2.

III. CONCLUSION

The production of methanol via an electrochemical process is found to be viable by considering a range of different rutile $MeO_2(110)$ surfaces and $AB_2C_2O_2$ MXenes as electrocatalysts. By varying the applied potential it is possible to tune the reactivity of an oxygen atom adsorbed on the surface. This is shown to correlate with the activation energy required to activate methane. A higher applied potential results in a lower energy barrier to overcome. In order to acquire sufficient selectivity the applied potential should not allow the full OER reaction to proceed and methanol needs to be harvested before being fully reduced to CO_2 and H_2O. Promising candidates are rutile IrO_2, $MoHf_2C_2O_2$ and $TaHf_2C_2O_2$. Only a small subset of MXenes have been studied here, but the performance warrants a larger study. By varying the number of layers, types of transition metal atoms and surface termination the combinations are almost endless and it should be possible to find optimal candidates just from calculations of the binding energy of O and OH.

IV. COMPUTATIONAL METHODS

All calculations were performed using the Grid-based Projector Augmented Wave (GPAW) program[17,18] with the Atomic Simulation Environment (ASE) interface[19].
For the MeO_2 structures: Exchange and correlations (XC) were modelled on the Generalized Gradient Approximation (GGA) level, using the XC-functional, BEEF-vdW[20], and grid spacing of $h = 0.18$ Å. The unit cell of the rutile (110) metal oxide slabs consisted of four tri-layers, in total 24 metal atoms and 48 oxygen atoms corresponding to a (1 x 3) surface unit cell. A k-point mesh of (3 x 3 x 1) was used to sample the Brilloin zone.
For the MXenes: The PBE[21] XC-functional was used and a plane wave cutoff of 800 eV employed. The 2×2 unit cell was used, resulting in 28 atoms. A k-point mesh of $(9 \times 9 \times 1)$ was necessary.
The unit cell of each material was optimised using the UnitCellFilter method available in the ASE package[19]. To model the transition state and obtain activation energies the Nudged Elastic Band method with the climbing image (CI-NEB) variation as implemented in ASE was used[21,22]. The Gibbs free energy of the OER reactions are calculated according to the methods described in 23.

ACKNOWLEDGMENTS

We acknowledge support from the Danish Council for Independent Research's Sapere Aude Program, Grant No. 11-1051390. The Center for Nanostructured Graphene is sponsored by the Danish National Research Foundation, Project DNRF58.

[1] The World Bank. Global Gas Flaring Reduction Partnership: Improving Energy Efficiency and Mitigating Impact on Climate Change (2011).

[2] US Energy Information Administration http://www.eia.gov/naturalgas.

[3] International Energy Agency. Key World Statistics (2014).

[4] P. Forster, V. Ramaswamy, P. Artaxo, T. Berntsen, R. Betts, D. Fahey, J. Haywood, J. Lean, D. Lowe, G. Myhre, J. Nganga, M. S. R. P. G. Raga, and R. V. Dorland, Changes in Atmospheric Constituents and in Radiative Forcing (2007).

[5] M. H. Ab Rahim, M. M. Forde, C. Hammond, R. L. Jenkins, N. Dimitratos, J. A. Lopez-Sanchez, A. F. Carley, S. H. Taylor, D. J. Willock, and G. J. Hutchings, TOPICS IN CATALYSIS **56** (2013).

[6] R. Palkovits, M. Antonietti, P. Kuhn, A. Thomas, and F. Schüth, Angewandte Chemie International Edition **48**, 6909 (2009).

[7] G. A. Olah, Angewandte Chemie International Edition **52**, 104 (2013).

[8] K. W. Frese, Langmuir **7**, 13 (1991), http://dx.doi.org/10.1021/la00049a004.

[9] A. Tomita, J. Nakajima, and T. Hibino, Angewandte Chemie International Edition **47**, 1462 (2008).

[10] G. Psofogiannakis, A. St-Amant, and M. Ternan, The Journal of Physical Chemistry B **110**, 24593 (2006), pMID: 17134220, http://dx.doi.org/10.1021/jp061559+.

[11] B. Lee and T. Hibino, Journal of Catalysis **279**, 233 (2011).

[12] Y. E. Seidel, A. Schneider, Z. Jusys, B. Wickman, B. Kasemo, and R. J. Behm, Langmuir **26**, 3569 (2010), pMID: 19968240, http://dx.doi.org/10.1021/la902962g.

[13] M. Naguib, J. Come, B. Dyatkin, V. Presser, P.-L. Taberna, P. Simon, M. W. Barsoum, and Y. Gogotsi, Electrochemistry Communications **16**, 61 (2012).

[14] M. Naguib, M. Kurtoglu, V. Presser, J. Lu, J. Niu, M. Heon, L. Hultman, Y. Gogotsi, and M. W. Barsoum, Advanced Materials **23**, 4248 (2011).

[15] M. Pandey and K. S. Thygesen, The Journal of Physical Chemistry C **121**, 13593 (2017), http://dx.doi.org/10.1021/acs.jpcc.7b05270.

[16] A. Latimer, H. Aljama, A. Kakekhani, J. S. Yoo, A. Kulkarni, C. Tsai, M. Garcia-Melchor, F. Abild-Pedersen, and J. K. Nørskov, Physical Chemistry Chemical Physics **19**, 3575 (2017).

[17] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, and K. W. Jacobsen, Journal of Physics: Condensed Matter **22**, 253202 (2010).

[18] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, Phys. Rev. B **71**, 035109 (2005).

[19] S. Bahn and K. Jacobsen, Computing in Science and Engineering **4**, 56 (2002), copyright: 2002 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

[20] J. Wellendorff, K. T. Lundgaard, A. Møgelhøj, V. Petzold, D. D. Landis, J. K. Nørskov, T. Bligaard, and K. W. Jacobsen, Phys. Rev. B **85**, 235149 (2012).

[21] H. JÓNSSON, G. MILLS, and K. W. JACOBSEN, "Nudged elastic band method for finding minimum energy paths of transitions," in *Classical and Quantum Dynamics in Condensed Phase Simulations*, pp. 385–404.

[22] G. Henkelman, B. P. Uberuaga, and H. Jónsson, The Journal of Chemical Physics **113**, 9901 (2000), http://dx.doi.org/10.1063/1.1329672.

[23] J. Rossmeisl, Z.-W. Qu, H. Zhu, G.-J. Kroes, and J. Nørskov, Journal of Electroanalytical Chemistry **607**, 83 (2007).

Paper III

Efficient many-body calculations of two-dimensional materials using exact limits for the screened potential: Band gaps of MoS_2, h-BN and phosphorene

Filip A. Rasmussen, Per S. Schmidt, Kirsten T. Winther and Kristian S. Thygesen

Phys. Rev. B. 94, 155406 (2016).

Efficient many-body calculations of 2D materials using exact limits for the screened potential: Band gaps of MoS_2, hBN and phosphorene

Filip A. Rasmussen, Per S. Schmidt, Kirsten T. Winther, and Kristian S. Thygesen
Center for Atomic-scale Materials Design (CAMD),
Department of Physics and Center for Nanostructured Graphene (CNG),
Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark
(Dated: September 9, 2016)

Calculating the quasiparticle (QP) band structure of two-dimensional (2D) materials within the GW self-energy approximation has proven to be a rather demanding computational task. The main reason is the strong $\mathbf{q}$-dependence of the 2D dielectric function around $\mathbf{q} = \mathbf{0}$ that calls for a much denser sampling of the Brillouin zone than is necessary for similar 3D solids. Here we use an analytical expression for the small $\mathbf{q}$-limit of the 2D response function to perform the BZ integral over the critical region around $\mathbf{q} = \mathbf{0}$. This drastically reduces the requirements on the $\mathbf{q}$-point mesh and implies a significant computational speed-up. For example, in the case of monolayer MoS_2, convergence of the G_0W_0 band gap to within $\sim 0.1\,\text{eV}$ is achieved with 12×12 $\mathbf{q}$-points rather than the 36×36 mesh required with discrete BZ sampling techniques. We perform a critical assessment of the band gap of the three prototypical 2D semiconductors MoS_2, hBN and phosphorene including the effect of self-consistency at the GW_0 level. The method is implemented in the open source GPAW code.

PACS numbers: 71.15.Dx, 71.15.Qe, 73.22.-f

I. INTRODUCTION

The past few years have witnessed an explosion in research on atomically thin two-dimensional (2D) materials. Of particular interest are the 2D semiconductors including the family of transition metal dichalcogenides, which have been found to exhibit a number of unique opto-electronic properties.[1–7] For understanding and predicting these properties the electronic band structure of the material is of fundamental importance. The GW method,[8,9] introduced by Hedin[10] in 1965 and first applied to real solids in an ab-initio framework by Hybertsen and Louie[11] and Godby, Sham, and Schlüter,[12] has become the "gold standard" for calculating quasi-particle (QP) band structures. Over the years its performance has been thoroughly established for bulk materials[13–15] and more recently also for molecules.[16–19] In comparison, critical assessments of the accuracy and numerical convergence of GW calculations for 2D materials are rather scarce.[20–23] Nevertheless these studies have shown that (i) a truncated Coulomb interaction can be used to avoid long range screening between periodically repeated layers which reduces the QP band gap, and (ii) when a truncated Coulomb interaction is used, the convergence of the GW calculation with respect to the number of $\mathbf{k}$-points becomes much slower than is the case for similar bulk systems.

The slow $\mathbf{k}$-point convergence of the GW band structure is directly related to the nature of electronic screening in 2D which is qualitatively different from the well known 3D case.[24,25] Specifically, while the dielectric function, $\varepsilon(\mathbf{q})$, of bulk semiconductors is approximately constant for small wave vectors, the dielectric function of a 2D semiconductor varies sharply as $\mathbf{q} \to \mathbf{0}$.[20,21] As a consequence, the number of $\mathbf{q}$-points required to obtain a proper sampling of the screened interaction $W(\mathbf{q})$ over the Brillouin zone (BZ) is much higher for the 2D material than what would be anticipated from the 3D case. For example, the band gap of bulk MoS_2 is converged to within $\sim 0.1\,\text{eV}$ with an in-plane $\mathbf{k}$-point grid of 12×12 while the same accuracy for monolayer MoS_2 requires a grid of 36×36 when standard BZ sampling schemes are applied.

Importantly, supercell calculations only describe the characteristic features of screening in 2D materials correctly when the unphysical screening between the periodically repeated layers is removed, e.g. using a truncated Coulomb interaction. Without this, the dielectric function does not approach unity for $\mathbf{q} \to 0$ and the band gap can be significantly underestimated (by around $0.5\,\text{eV}$ for MoS_2 with $10\,\text{Å}$ vacuum[21]) as a result of over screening. Since in this case, the screening is really intermediate between 3D and 2D, the GW gap shows faster convergence with k-points than is observed using a truncated Coulomb interaction. This is presumably the reason why most earlier GW calculations for 2D semi-conductors have been performed with k-point grids ranging from 6×6 to 15×15 which are much too coarse for describing the truly isolated 2D material.

Here we show that the slow $\mathbf{k}$-point convergence of the GW self-energy in 2D materials can be alleviated by performing the BZ integral of $W(\mathbf{q})$ analytically in the critical region around $\mathbf{q} = \mathbf{0}$ where $\varepsilon(\mathbf{q})$ varies most strongly. The analytical expression for $W(\mathbf{q})$ is obtained from a lowest order expansion in $\mathbf{q}$ of the head, $\chi^0_{\mathbf{00}}(\mathbf{q})$, and wings, $\chi^0_{\mathbf{0G}}(\mathbf{q})$, of the non-interacting density response function. This simple scheme reduces the required number of $\mathbf{q}$-points by an order of magnitude without loss of accuracy.

II. THE GW SELF-ENERGY

We split the GW self-energy into the exchange and correlation part, respectively. The former does not present particular problems in 2D materials and is calculated using a Wigner-Seitz truncated Coulomb interaction as described elsewhere.[26] In a plane wave expansion the correlation part of the self-energy, evaluated for an electronic state $|n\mathbf{k}\rangle$, takes the general form[27]

$$\langle n\mathbf{k}|\Sigma^c(\omega)|n\mathbf{k}\rangle = \frac{1}{(2\pi)^3}\int_{\mathrm{BZ}} d\mathbf{q}\sum_{\mathbf{G}\mathbf{G}'}\frac{i}{2\pi}\int_{-\infty}^{\infty}d\omega'\overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega') \times\sum_m \frac{[\rho_{n\mathbf{k}}^{m\mathbf{k}+\mathbf{q}}(\mathbf{G})][\rho_{n\mathbf{k}}^{m\mathbf{k}+\mathbf{q}}(\mathbf{G}')]^*}{\omega+\omega'-\epsilon_{m\mathbf{k}+\mathbf{q}}+i\eta\,\mathrm{sgn}(\epsilon_{m\mathbf{k}+\mathbf{q}}-\mu)}, \tag{1}$$

where m runs over all electronic bands, $\epsilon_{m\mathbf{k}+\mathbf{q}}$ are single particle energies, μ is the chemical potential and η is a broadening parameter set to 0.1 eV. The pair densities are defined as $\rho_{n\mathbf{k}}^{m\mathbf{k}+\mathbf{q}}(\mathbf{G}) = \langle n\mathbf{k}|e^{i(\mathbf{G}+\mathbf{q})\cdot\mathbf{r}}|m\mathbf{k}+\mathbf{q}\rangle$, and $\overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)$ is the correlation part of the dynamical screened potential given by

$$\overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) = v_{\mathbf{G}}(\mathbf{q})\left[\varepsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)-\delta_{\mathbf{G}\mathbf{G}'}\right]v_{\mathbf{G}'}(\mathbf{q}), \tag{2}$$

where $v_{\mathbf{G}}(\mathbf{q}) = 4\pi/|\mathbf{G}+\mathbf{q}|^2$ is the Coulomb interaction. In most implementations the BZ integral is evaluated numerically with a standard quadrature method using a regular $\mathbf{q}$-point grid matching the $\mathbf{k}$-point grid of the ground state DFT calculation. Since the screened potential, Eq. (2), diverges for $\mathbf{G}=0$, $\mathbf{q}=\mathbf{0}$ (for bulk materials) this point is generally handled separately, so the integral may be written

$$\int_{\mathrm{BZ}} d\mathbf{q}\mathcal{S}(\mathbf{q},\omega) \to \frac{\Omega}{N_{\mathbf{q}}}\sum_{\mathbf{q}_n\neq\mathbf{0}}\mathcal{S}(\mathbf{q}_n,\omega)+\int_{\Omega_0}d\mathbf{q}\mathcal{S}(\mathbf{q},\omega), \tag{3}$$

where $\mathcal{S}(\mathbf{q},\omega)$ denotes the entire integrand, Ω is the volume of the BZ, $N_{\mathbf{q}}$ is the total number of $\mathbf{q}$-points in the grid and Ω_0 denotes a small region around $\mathbf{q}=\mathbf{0}$. For bulk systems Ω_0 is normally defined as a sphere centered at $\mathbf{q}=\mathbf{0}$. For a 2D material, the BZ integral in Eq. (1) reduces to a 2D integral with in-plane sampling of $\mathbf{q}$, and Ω_0 represents an area. For the Ω_0 region we assume that all the terms in the integrand but the screened potential is constant in $\mathbf{q}$ and therefore take the $\mathbf{q}=0$ value. As a result, the integral can be written on the following form

$$\int_{\Omega_0} d\mathbf{q}\overline{W}(\mathbf{q},\omega)\mathcal{G}(\mathbf{q},\omega) \approx \mathcal{G}(\mathbf{0},\omega)\int_{\Omega_0}d\mathbf{q}\overline{W}(\mathbf{q},\omega)$$

We now focus on how to calculate the contribution to the integral around the special point $\mathbf{q}=\mathbf{0}$ in the 3D versus the 2D case.

Within the random phase approximation (RPA) the dielectric matrix is given by

$$\varepsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) = \delta_{\mathbf{G}\mathbf{G}'} - v_{\mathbf{G}}(\mathbf{q})\chi^0_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega). \tag{4}$$

with the non-interacting response function

$$\chi^0_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) = \frac{1}{V}\sum_{\mathbf{k}}^{\mathrm{BZ}}\sum_{n,n'}\frac{f_{n\mathbf{k}}-f_{n'\mathbf{k}+\mathbf{q}}}{\omega+\epsilon_{n\mathbf{k}}-\epsilon_{n'\mathbf{k}+\mathbf{q}}+i\eta}\times [\rho_{n\mathbf{k}}^{n',\mathbf{k}+\mathbf{q}}(\mathbf{G})][\rho_{n\mathbf{k}}^{n'\mathbf{k}+\mathbf{q}}(\mathbf{G}')]^*$$

where $f_{n\mathbf{k}}$ are the occupation factors and V the volume of the unit cell. For a solid with a finite band gap it can be shown that the head of the non-interacting response function $\chi^0_{\mathbf{0}\mathbf{0}}(\mathbf{q},\omega) \propto q^2$ for small q.[28] Since $v_{\mathbf{0}}(\mathbf{q}) = 4\pi/q^2$ it follows that in 3D the head of the dielectric function $\varepsilon_{\mathbf{0}\mathbf{0}}(\mathbf{q},\omega)$ converges to a finite value > 1 when $q\to 0$. Moreover, this value is typically a reasonable approximation to $\varepsilon_{\mathbf{0}\mathbf{0}}(\mathbf{q},\omega)$ in a relatively large region around $\mathbf{q}=\mathbf{0}$. This means that in the BZ integration in Eq. (1) around the singular point $\mathbf{G}=\mathbf{G}'=\mathbf{q}=\mathbf{0}$ all factors, except $4\pi/q^2$, can be assumed to be constant and the integral can be performed analytically over a sphere centred at $\mathbf{q}=\mathbf{0}$[11] or numerically on a fine real space grid.[29]

For GW calculations of 2D materials performed with periodic boundary conditions in the out-of-plane direction, the direct use of Eq. (1) leads to significant overscreening due to the interaction between the repeated images.[21] One way of dealing with this is to subtract the artificial interlayer contribution calculated from a classical electrostatic model.[30] A more direct way of avoiding the spurious interactions is to truncate the Coulomb potential in the direction perpendicular to the layers. Thus in Eqs. (4) and (2), $v_{\mathbf{G}}(\mathbf{q})$ should be replaced by[20,31]

$$v^{\mathrm{2D}}_{\mathbf{G}}(\mathbf{q}_\parallel) = \frac{4\pi}{|\mathbf{q}_\parallel+\mathbf{G}|^2}\left[1-e^{-|\mathbf{q}_\parallel+\mathbf{G}_\parallel|L/2}\cos(G_zL/2)\right], \tag{5}$$

where $q_z=0$ such that only in-plane $\mathbf{q}$ are included in the sampling of the BZ. L is the length of the unit cell in the non-periodic direction and the truncation distance is set to $L/2$, which simplifies the expression. In the long wavelength limit, $\mathbf{G}=\mathbf{0}$, $q_z=0$, $\mathbf{q}_\parallel\to\mathbf{0}$, the truncated interaction becomes $v^{\mathrm{2D}}_{\mathbf{0}}(q_\parallel)\approx\frac{2\pi L}{q_\parallel}$. We see that the $\mathbf{q}=\mathbf{0}$ divergence in the truncated Coulomb potential is reduced by a power of $\mathbf{q}$ compared to that of the full Coulomb interaction. As will be shown in the following this has important consequences for the form of the screened interaction. However, before presenting the form of the screened interaction of a 2D semiconductor evaluated using the full expression for the response function and truncated Coulomb interaction, it is instructive to consider a simplified description of the 2D material.

Let us consider a strict 2D model of a homogeneous and isotropic semiconductor. In the small $\mathbf{q}$ limit, the density response function takes the form $\chi_0(\mathbf{q}) = -\alpha_{\mathrm{2D}}q^2$ where α_{2D} is the 2D polarizability.[25] Using that the 2D Fourier transform of $1/r$ equals $2\pi/q$, the leading order of the dielectric function becomes

$$\varepsilon^{2D}(\mathbf{q}) \approx 1+2\pi\alpha_{\mathrm{2D}}q. \tag{6}$$

Some examples of *macroscopic* dielectric functions for a representative set of 2D semiconductors are shown

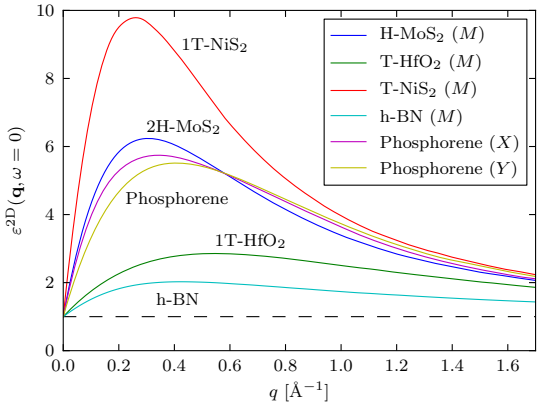


FIG. 1. (Color online) Static macroscopic dielectric functions of a representative set of 2D semiconductors as a function of $\mathbf{q}$ along the $\Gamma \to M$ direction for the hexagonal structures and along the path from Γ to X or Y in the case of phosphorene.

in Figure 1 (see Ref. 21 for a precise definition of this quantity). The linear form (6) is clearly observed in the small $\mathbf{q}$ regime. Importantly, if we use the same strategy for evaluating the BZ integral in Eq. (1) as in 3D, i.e. assuming $\epsilon^{-1}(\mathbf{q})$ to be a slowly varying function and evaluating it on the discrete $\mathbf{q}$-point grid, we obtain zero contribution for the $\mathbf{q} = 0$ term because $1/\epsilon^{2D} - 1 = 0$ for $\mathbf{q} = 0$, see Eq. (2). On the other hand, it is clear that the screened interaction takes the form $\overline{W}^{2D}(\mathbf{q}) = -4\pi^2\alpha_{2D}/(1 + 2\pi\alpha_{2D}q)$ for small $\mathbf{q}$. In particular, $\overline{W}^{2D}(\mathbf{q})$ takes a finite value for $\mathbf{q} = 0$ which is qualitatively different from the 3D case where $\overline{W}(\mathbf{q})$ diverges for $\mathbf{q} \to \mathbf{0}$.

In Appendix A we show, following an analysis similar to that of Ref. 32 adapted to the case of a truncated Coulomb interaction, that for a general non-isotropic 2D material, the small $\mathbf{q}_\parallel$ limit of the head of the screened potential takes the form

$$\overline{W}_{\mathbf{00}}(\mathbf{q}_\parallel) = -\left(\frac{4\pi(1-e^{-|\mathbf{q}_\parallel|L/2})}{|\mathbf{q}_\parallel|}\right)^2 \times \frac{\hat{\mathbf{q}}_\parallel \cdot \mathsf{A}\hat{\mathbf{q}}_\parallel}{1+4\pi(1-e^{-|\mathbf{q}_\parallel|L/2})\hat{\mathbf{q}}_\parallel \cdot \mathsf{A}\hat{\mathbf{q}}_\parallel}, \tag{7}$$

where $\hat{\mathbf{q}}_\parallel = \mathbf{q}_\parallel/|\mathbf{q}_\parallel|$ and A is a second rank tensor which also depends on the frequency. We see that we have $\overline{W}_{\mathbf{00}}(\mathbf{q}_\parallel = \mathbf{0}) = -(2\pi L)^2\hat{\mathbf{q}}_\parallel \cdot \mathsf{A}\hat{\mathbf{q}}_\parallel$. The expression $\hat{\mathbf{q}}_\parallel \cdot \mathsf{A}\hat{\mathbf{q}}_\parallel$ is closely related to the slope of the dielectric function and the 2D polarizability but includes local field effects. In addition to Eq. (7) there are similar expressions for the wings and body of the screened interaction, see Eq.(A25) to Eq.(A28), which are all included in the subsequent calculations. These expressions must be integrated over the Ω_0-region, that we now define as the primitive cell in the 2D BZ that surrounds the $\mathbf{q}_\parallel = \mathbf{0}$ point. The expression is simplified to one that can be integrated analytically as shown in Appendix A.

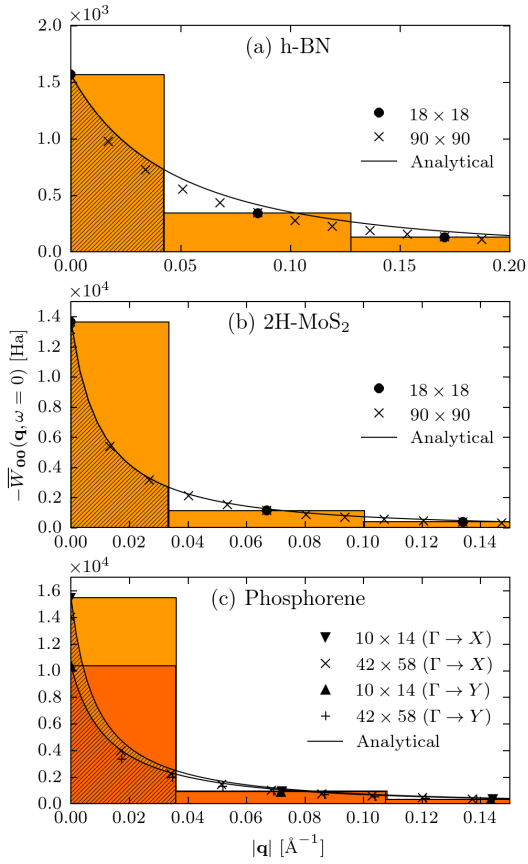


FIG. 2. (Color online) The head of the static component of the screened potential (subtracted the bare interaction) of monolayer a) h-BN b) MoS_2 and c) phosphorene as a function of $\mathbf{q}$ along the $\Gamma \to M$ direction or $\Gamma \to X$ and $\Gamma \to Y$ in the case of phosphorene. The crosses are the numerical values obtained on a fine $\mathbf{q}$-point grid while the circles or triangles represent the values obtained on a coarse $\mathbf{q}$-point grid. The bars represent a simple numerical approximation to the BZ integral of $\overline{W}_{00}(\mathbf{q})$ performed on the coarse $\mathbf{q}$-point grid. The value of the screened potential for $\mathbf{q} = \mathbf{0}$ is set to the analytical result Eq. (7). The full curve represents the analytical small $\mathbf{q}$ approximation, Eq. (7), and the hatched area shows its contribution to the integral.

The full expression for $\overline{W}$ in Eq. (7) is therefore evaluated numerically on a discrete sub-grid, constructed as a Monkhorst-Pack grid within Ω_0, and the simplified expression in Eq. (A31) is only used for $\mathbf{q}_\parallel = \mathbf{0}$ on the sub-grid. The limit of the integral is now given by the radius r_{Ω_0}, defined as $\pi r_{\Omega_0}^2 = \Omega_0/N_{q_0}$, where N_{q_0} is the

number of grid points in the sub-grid. This approach ensures a smooth evaluation of $\overline{W}$, that converges fast with N_{q_0}. It is found to be necessary to have $N_{q_0} \approx 100$ when $\mathbf{q}_\parallel = \mathbf{0}$ is evaluated using Eq. (A31) for both iso- and anisotropic materials where as $N_{\mathbf{q_0}} \approx 10^5$ is needed if the analytical correction at $\mathbf{q}_\parallel = \mathbf{0}$ is omitted.

A. Results

To investigate how this method performs we have carried out test calculations for the three monolayers h-BN, MoS_2 and phosphorene, which have quite different dielectric functions as seen on Figure 1. h-BN is a large gap dielectric with low screening ability leading to a small slope of the dielectric function at $\mathbf{q} = \mathbf{0}$, while MoS_2 has a larger dielectric function and quite steep slope at $\mathbf{q} = \mathbf{0}$. Phosphorene has a dielectric function similar to MoS_2 in size and steepness but is anisotropic with slopes varying by $\sim 40\%$ between the two high symmetry directions, $\Gamma \to X$ and $\Gamma \to Y$.

All the calculations were performed using the GPAW electronic structure code.[33–35] The structures used in the present calculations are relaxed with DFT using the PBE exchange-correlation (xc) functional.[36] The resulting lattice constant for h-BN is 2.504 Å, the in-plane lattice constant for MoS_2 is 3.184 Å with a S-S distance of 3.127 Å. For phosphorene the in-plane unit cell is 4.630 Å by 3.306 Å, the in-plane P-P-P angle is 95.8° and the layer thickness is 2.110 Å. A convergence test with respect to the amount of vacuum between repeating periodic images was carried out and 15 Å of vacuum was necessary for h-BN and Phosphorene where as only 10 Å was needed for MoS_2. The PBE eigenvalues and wavefunctions were calculated with a plane wave basis cut-off energy of 600 eV and used as input in the GW calculations. For the initial investigation of the $\mathbf{q}$-point convergence, the dielectric function and the correlation self-energy were calculated using a cutoff of 50 eV. This cutoff is insufficient to ensure properly converged quasiparticle energies, but it is adequate to describe the trends related to the improved $\mathbf{q}$-point sampling relevant for this study. The following fully converged calculations were carried out using a $1/N_{\mathrm{pw}}$ extrapolation to the complete basis set limit using cutoff energies of up to 200 eV.[37–39]

In Figure 2 we compare the analytical small $\mathbf{q}$ expression, Eq. (7), for the head of the screened potential $\overline{W}_{\mathbf{00}}(\mathbf{q})$ with the numerical values obtained using a fine and coarse $\mathbf{q}$-point sampling. In all the cases the $\mathbf{q} = \mathbf{0}$ value has been set to the analytical value. It is evident that the screened potential falls off quickly and thus for a coarse $\mathbf{q}$-point sampling the $\mathbf{q} = \mathbf{0}$ contribution to the integral is by far the largest and should therefore not be neglected. Similarly, using only the exact value in $\mathbf{q} = \mathbf{0}$ could also pose a problem as the contribution will be grossly overestimated due to the convex nature of potential. We note that the analytical expression follows the numerical results quite closely and is even accurate far away from the Γ-point – for MoS_2 we have an almost perfect agreement for the points shown. Thus using the analytical limit within the region around $\mathbf{q} = \mathbf{0}$ is reasonable. We notice that the anisotropy of phosphorene makes $\overline{W}_{\mathbf{00}}(\mathbf{q})$ ill-defined at $\mathbf{q} = \mathbf{0}$ (different limit values depending on the direction of $\mathbf{q}$). For larger $\mathbf{q}$ the dielectric anisotropy becomes negligible. However, because of the relatively large weight of the $\mathbf{q} = \mathbf{0}$ contribution to the BZ integral, the anisotropy should be taken into account for accurate GW calculations.

We note that a similar approach to the treatment of the $\mathbf{q} = \mathbf{0}$ term of the screened potential was suggested in Ref. 20. That particular method was based on fitting to an empirical expression for $\varepsilon(q)$ calculated from the value at a small but finite $\mathbf{q}$. The method outlined here is different in that the analytical expression for $\overline{W}(\mathbf{q})$ is obtained from a lowest order expansion of the head, $\chi^0_{\mathbf{00}}(q)$, and wings, $\chi^0_{\mathbf{0G}}(q)$, of the non-interacting density response function[40] and thus can be obtained without fitting or using empirical parameters. This also ensures that the effect of in-plane dielectric anisotropy is explicitly included.

In Fig. 3 we show the minimum QP band gap of monolayer h-BN, MoS_2 and phosphorene as a function of $1/N_{\mathbf{k}}$ where $N_{\mathbf{k}}$ is the total number of $\mathbf{k}$-points in the BZ sampling (the $\mathbf{q}$ point grid for the GW integration is the same as the $\mathbf{k}$-point grid used in DFT). We compare the results obtained using two methods: i) neglecting the $\mathbf{q} = \mathbf{0}$ contribution to head and wings of the screened potential and ii) evaluating Eq. (7) as described. It is clear that method i) in all cases underestimates the correlation self-energy due to the underestimation of the screening; In order to get the band gap converged to within ~ 0.1 eV one would have to use a $\mathbf{k}$-point sampling of minimum $36 \times 36 \times 1$ for h-BN, $36 \times 36 \times 1$ for MoS_2 and $22 \times 30 \times 1$ for phosphorene. We also note that for large $\mathbf{k}$-point grids the band gaps using this method converge approximately as $1/N_{\mathbf{k}}$ as the missing contribution is almost proportional to the area of the $\mathbf{q} = \mathbf{0}$ region. Clearly, the latter approach varies significantly less with the $\mathbf{k}$-point grid and in fact the gap is converged to within 0.2 eV already for a $\mathbf{k}$-point grid in the order of $6 \times 6 \times 1$ and to within ~ 0.1 eV with a $12 \times 12 \times 1$ grid (in the worst case). We have performed test calculations for other 2D semiconductors and obtained similar conclusions although the number of $\mathbf{k}$-points required to reach convergence within 0.1 eV following the conventional approach ($\mathbf{q} = \mathbf{0}$ term neglected) is somewhat system dependent; materials with efficient screening, e.g. MoS_2 and NiS_2, require larger $\mathbf{k}$-point grids than materials with poor screening, e.g. h-BN and HfO_2 (see Fig. 1).

To obtain converged band gaps it is necessary to use a unit cell with enough vacuum between repeated layers to avoid an artificial interaction. This is true even when a truncated Coulomb interaction is used as the finite vacuum affects wave functions and energies, in particular for higher lying unbound states. As the amount of vacuum is increased, the Brillouin zone shrinks and the analyti-

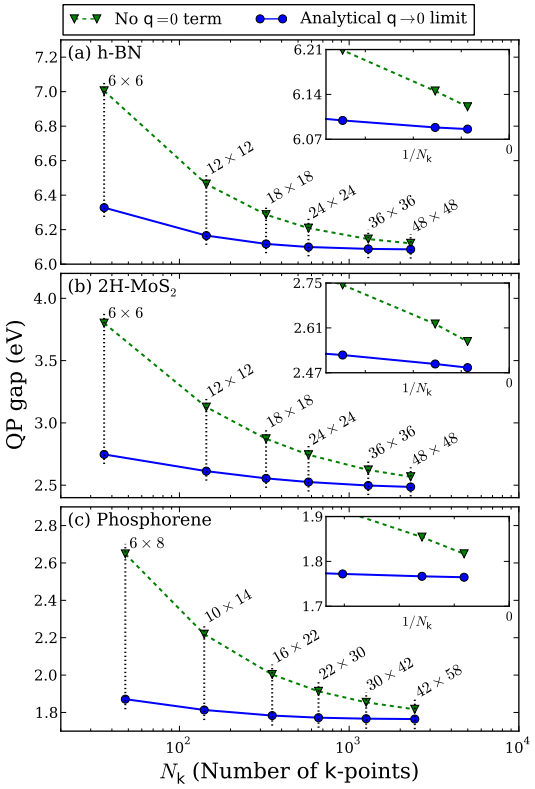


FIG. 3. (Color online) The G_0W_0 quasi-particle band gap of monolayer (a) 2H-MoS_2 (b) h-BN and (c) phosphorene, calculated using two different treatments of the $\mathbf{q} = \mathbf{0}$ term in Eq. (1). The dashed (green) line shows the contribution obtained when the head and wing elements of the $\mathbf{q} = \mathbf{0}$ term are neglected corresponding to the standard treatment used for 3D systems. The solid (blue) line shows the contribution obtained when using the analytical results, Eq. (7), to perform the integral over the $\mathbf{q} = \mathbf{0}$ element. The insets shows the results for the largest $\mathbf{k}$-point grids on a reversed linear scale in $1/N_{\mathbf{k}}$. Notice the zero point is at the right side of the x-axis.

cal correction around $\mathbf{q} = 0$, applied only for $\mathbf{G} = 0$, has smaller weight. This means a slower convergence with respect to in-plane $\mathbf{k}$-points. This is shown in Fig. 4 for MoS_2, where it is clear that the correction is less effective for larger vacuum. The calculations converge toward the same value indicating that for MoS_2 10 Å of vacuum is sufficient. The most efficient procedure to obtain converged band gaps is therefore to first converge the amount of vacuum at a low $\mathbf{k}$-point sampling without applying the correction and then afterwards converge the $\mathbf{k}$-point sampling with the correction at the given vacuum.

In Table I we report the converged values for the quasi-

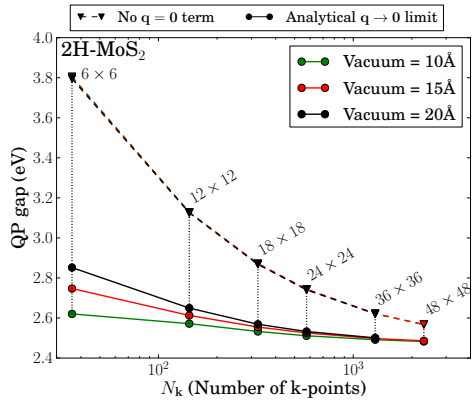


FIG. 4. (Color online) The band gap of monolayer 2H-MoS_2 calculated with different amounts of vacuum between repeated layers. The solid and dashed lines are with and without the $\mathbf{q} = 0$ correction, respectively. As the vacuum is increased, the weight of the correction is decreased and it is necessary to use denser in-plane $\mathbf{k}$-point sampling to achieve convergence.

particle band gaps. For h-BN the band gap is indirect between the K- and Γ-point, for MoS_2 and phosphorene it is direct at the K- and Γ-point respectively. For these calculations we used $18\times18\times1$ $\mathbf{k}$-points for h-BN, $18\times18\times1$ $\mathbf{k}$-points for MoS_2 and $10 \times 14 \times 1$ for phosphorene with the analytical integration of $W(\mathbf{q})$ around $\mathbf{q} = \mathbf{0}$. According to Fig. 3 this is sufficient to ensure convergence to within 0.05 eV. We note that spin-orbit interactions are not included in the reported values. Inclusion of spin-orbit interactions split the valence band of MoS_2 at the K point by 0.15 eV thereby lowering the QP gap by around 0.07 eV.[23,41] Spin-orbit interactions have negligible effect for h-BN and phosphorene.

For MoS_2 the converged G_0W_0@PBE band gap of 2.54 eV agrees well with our previously reported value of 2.48 eV (with spin-orbit coupling) obtained using a Wigner-Seitz truncated Coulomb interaction and $30 \times 30\times1$ $\mathbf{k}$-points.[23] Other reported gaps range from 2.40 eV to 2.82 eV.[42–47] However, these calculations were performed i) without the use of a truncated Coulomb interaction and including 15-25 Å vacuum, ii) employing relatively small $\mathbf{k}$-point grids of $6 \times 6 \times 1$ to $18 \times 18 \times 1$, and iii) using different in-plane lattice constants varying between 3.15 and 3.19 Å. These different settings can affect the band gap by as much as 0.5 eV,[35] and therefore we refrain from providing detailed comparison of our result to these earlier calculations. An overview of previous GW results for MoS_2 can be found in Ref. 35.

In Ref. 22 a G_0W_0@LDA band gap for MoS_2 of 2.70 eV is reported using a truncated Coulomb interaction and a calculation of the screened potential at $\mathbf{q} = \mathbf{0}$ based on the method in Ref. 20. In that study, the lattice constant of MoS_2 was 3.15 Å. With this lattice constant we

	Transition	DFT-PBE	G_0W_0@PBE	GW_0@PBE
h-BN	$K \to \Gamma$	4.64	7.06	7.49
	$K \to K$	4.72	7.80	8.25
2H-MoS_2	$K \to K$	1.65	2.54	2.65
Phosphorene	$\Gamma \to \Gamma$	0.90	2.03	2.29

TABLE I. Band gaps in eV calculated with DFT-PBE, G_0W_0@PBE and GW_0@PBE using the PBE-relaxed structures. The GW calculations were performed using analytic integration of $\overline{W}(\mathbf{q})$ around $\mathbf{q} = \mathbf{0}$ without including spin-orbit interactions. 10 Å of vacuum was used for MoS_2 and 15 Å for h-BN and phosphorene. The following $\mathbf{k}$-point grids were used; h-BN: $18 \times 18 \times 1$, 2H-MoS2: $18 \times 18 \times 1$ and phosphorene: $10 \times 14 \times 1$.

obtain a gap of 2.64 eV, which is in fair agreement with Ref. 22. Our result is very close to the experimental value of 2.5 eV inferred from photo current spectroscopy.[48] Performing partially self-consistent GW_0 the band gap increases to 2.65 eV (2.58 eV including spin-orbit coupling).

For h-BN, we obtained a G_0W_0 band gap of 7.06 eV which increases to 7.49 eV with GW_0. In Ref. 49 the G_0W_0 band gap was calculated to be 7.40 eV. Instead of a truncated Coulomb interaction the band gap was extrapolated to infinite vacuum. The treatment of the $\mathbf{q} = 0$ term is not mentioned nor is the size of the $\mathbf{k}$-point grid. Despite the difference at the G_0W_0 level, they report a similar increase of the band gap of 0.4 eV when doing a GW_0 calculation.

For phosphorene we calculate a G_0W_0 band gap of 2.03 eV which agrees well with the previously reported value of 2.0 eV[50] using the method of Ref. 20. The band gap increases to 2.29 eV with GW_0.

III. CONCLUSION

In conclusion, we have discussed the connection between the form of the $\mathbf{q}$-dependent dielectric function of a 2D semiconductor and the slow $\mathbf{k}$-point convergence of the GW band structure. We have derived an analytical expression for the $\mathbf{q} \to \mathbf{0}$ limit of the screened potential of a semiconductor when a 2D truncation of the Coulomb potential is used. The method accounts for dielectric anisotropy in an approximate way (see discussion below Eq.(A29)) and does not rely on any additional parameters or fitting. Using this expression we have shown that convergence of the GW self-energy with respect to the size of the $\mathbf{k}$-point grid is drastically improved. For the specific case of monolayer MoS_2, we found that the use of the analytical form alone reduces the $\mathbf{k}$-point grid required to achieve convergence of the GW self-energy contribution to the band gap to within ~ 0.1 eV from around $36 \times 36 \times 1$ to $12 \times 12 \times 1$ – a reduction in the number of $\mathbf{k}$-points by an order of magnitude. This method may therefore enable future large-scale GW calculations for 2D materials without compromising accuracy.

ACKNOWLEDGMENTS

We acknowledge support from the Danish Council for Independent Research's Sapere Aude Program, Grant No. 11-1051390. The Center for Nanostructured Graphene is sponsored by the Danish National Research Foundation, Project DNRF58.

Appendix A: Calculation of the $\mathbf{q} \to \mathbf{0}$ limit of the screened potential

In the following we derive the analytical form of the screened potential, Eq. (2), for 2D materials in the limit $\mathbf{q}_\parallel \to \mathbf{0}$. We largely follow the approach of Ref. 32 where the same limit for bulk systems was considered. As explained in the main text we use a truncated Coulomb interaction of the form

$$v(\mathbf{r}_\parallel, z) = \frac{\theta(R - |z|)}{\sqrt{|\mathbf{r}_\parallel|^2 + z^2}}. \tag{A1}$$

Using this potential we effectively turn off interaction between electrons on different 2D layers of the supercell calculation. We choose R to be half the height of the unitcell, $R = L/2$, so that an electron in the center of the layer will not interact with electrons located in the periodically repeated layer. This means that the 2D truncated coulomb interaction of Eq. (16) in[20] reduces to

$$v^{\mathrm{2D}}_{\mathbf{G}}(\mathbf{q}_\parallel) = \frac{4\pi}{|\mathbf{q}_\parallel + \mathbf{G}|^2}\left[1 - e^{-|\mathbf{q}_\parallel + \mathbf{G}_\parallel| L/2}\cos(|G_z|L/2)\right], \tag{A2}$$

where only in-plane $\mathbf{q}$ is considered. We note that in the limit $L \to \infty$ it takes the usual 3D form, $v_{\mathbf{G}}(\mathbf{q}) = \frac{4\pi}{|\mathbf{q}+\mathbf{G}|^2}$. In the long wavelength limit it has the asymptotic behavior

$$v^{\mathrm{2D}}_{\mathbf{0}}(\mathbf{q}_z = 0, \mathbf{q}_\parallel \to \mathbf{0}) = \frac{2\pi L}{|\mathbf{q}_\parallel|}, \tag{A3}$$

diverging slower than the full Coulomb potential with profound consequences for the properties of 2D materials.

In the long wavelength limit $\mathbf{q} \to \mathbf{0}$ the non-interacting density response function or irreducible polarizability has the following behavior[40]

$$\chi^0_{\mathbf{00}'}(\mathbf{q} \to \mathbf{0}) = \mathbf{q} \cdot \mathsf{P}\mathbf{q} = |\mathbf{q}|^2 \hat{\mathbf{q}} \cdot \mathsf{P}\hat{\mathbf{q}} \tag{A4}$$

$$\chi^0_{\mathbf{G0}}(\mathbf{q} \to \mathbf{0}) = \mathbf{q} \cdot \mathbf{p}_{\mathbf{G}} = |\mathbf{q}|\hat{\mathbf{q}} \cdot \mathbf{p}_{\mathbf{G}}, , \tag{A5}$$

$$\chi^0_{\mathbf{0G}}(\mathbf{q} \to \mathbf{0}) = \mathbf{q} \cdot \mathbf{s}_{\mathbf{G}} = |\mathbf{q}|\hat{\mathbf{q}} \cdot \mathbf{s}_{\mathbf{G}}, \tag{A6}$$

where P is a second rank tensor, $\mathbf{p}_{\mathbf{G}}$ and $\mathbf{s}_{\mathbf{G}}$ are proper vectors and $\hat{\mathbf{q}} = \mathbf{q}/|\mathbf{q}|$. The density response function, and therefore also P and $\mathbf{p}_{\mathbf{G}}$, has a frequency dependence which here and through the rest of this section has been left out to simplify the notation. Within the random phase approximation the dielectric function is given by (schematically)

$$\varepsilon = 1 - v\chi^0. \tag{A7}$$

Due to technical reasons[13,51] it is easier to work with a similar symmetrized version given in Fourier space by

$$\tilde{\varepsilon}_{\mathbf{GG'}}(\mathbf{q}) = \delta_{\mathbf{GG'}} - \sqrt{v_{\mathbf{G}}(\mathbf{q})}\chi^0_{\mathbf{GG'}}(\mathbf{q})\sqrt{v_{\mathbf{G'}}(\mathbf{q})}. \quad \text{(A8)}$$

Inserting the Coulomb potential, Eq. (A2), and the expressions for the non-interacting response function Eqs. (A4)-(A6), the head and wings of the symmetrized dielectric function are

$$\begin{aligned}\tilde{\varepsilon}_{\mathbf{00}}(\mathbf{q}_\parallel \to \mathbf{0}) &= 1 - v_{\mathbf{0}}^{\mathrm{2D}}(\mathbf{q}_\parallel)|\mathbf{q}_\parallel|^2 \hat{\mathbf{q}}_\parallel \cdot \mathsf{P}\hat{\mathbf{q}}_\parallel \\ &= 1 - 4\pi(1 - e^{-|\mathbf{q}_\parallel| L/2})\hat{\mathbf{q}}_\parallel \cdot \mathsf{P}\hat{\mathbf{q}}_\parallel \quad \text{(A9)}\\ \tilde{\varepsilon}_{\mathbf{G0}}(\mathbf{q}_\parallel \to \mathbf{0}) &= -\sqrt{v_{\mathbf{G}}^{\mathrm{2D}}(\mathbf{0})}\sqrt{v_{\mathbf{0}}^{\mathrm{2D}}(\mathbf{q}_\parallel)}\hat{\mathbf{q}}_\parallel \cdot \mathbf{p}_{\mathbf{G}} \\ &= -\sqrt{v_{\mathbf{G}}^{\mathrm{2D}}(\mathbf{0})}\sqrt{4\pi(1 - e^{-|\mathbf{q}_\parallel| L/2})}\hat{\mathbf{q}}_\parallel \cdot \mathbf{p}_{\mathbf{G}} \quad \text{(A10)}\\ \tilde{\varepsilon}_{\mathbf{0G}}(\mathbf{q}_\parallel \to \mathbf{0}) &= -\sqrt{v_{\mathbf{G}}^{\mathrm{2D}}(\mathbf{0})}\sqrt{v_{\mathbf{0}}^{\mathrm{2D}}(\mathbf{q}_\parallel)}\hat{\mathbf{q}}_\parallel \cdot \mathbf{s}_{\mathbf{G}} \\ &= -\sqrt{v_{\mathbf{G}}^{\mathrm{2D}}(\mathbf{0})}\sqrt{4\pi(1 - e^{-|\mathbf{q}_\parallel| L/2})}\hat{\mathbf{q}}_\parallel \cdot \mathbf{s}_{\mathbf{G}}. \quad \text{(A11)}\end{aligned}$$

To determine the inverse dielectric function we write the dielectric function as a block matrix in the $\mathbf{G}, \mathbf{G'}$ components with head, wings and body of the form

$$\tilde{\varepsilon} = \begin{pmatrix} H & \mathbf{w}^\intercal \\ \mathbf{v} & \mathbf{B} \end{pmatrix} \quad \text{(A12)}$$

The inverse is then given by

$$\tilde{\varepsilon}^{-1} = \begin{pmatrix} (H - \mathbf{w}^\intercal \mathbf{B}^{-1}\mathbf{v})^{-1} & -(H - \mathbf{w}^\intercal \mathbf{B}^{-1}\mathbf{v})^{-1}\mathbf{w}^\intercal \mathbf{B}^{-1} \\ -\mathbf{B}^{-1}\mathbf{v}(H - \mathbf{w}^\intercal \mathbf{B}^{-1}\mathbf{v})^{-1} & \mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{v}(H - \mathbf{w}^\intercal \mathbf{B}^{-1}\mathbf{v})^{-1}\mathbf{w}^\intercal \mathbf{B}^{-1} \end{pmatrix} \quad \text{(A13)}$$

From this we see that

$$\begin{aligned}\tilde{\varepsilon}^{-1}_{\mathbf{00}} &= \left[\tilde{\varepsilon}_{\mathbf{00}} - \sum_{\mathbf{G},\mathbf{G'}\neq\mathbf{0}} \tilde{\varepsilon}_{\mathbf{0G}} B^{-1}_{\mathbf{GG'}} \tilde{\varepsilon}_{\mathbf{G'0}}\right]^{-1} \quad \text{(A14)}\\ \tilde{\varepsilon}^{-1}_{\mathbf{G0}} &= -\tilde{\varepsilon}^{-1}_{\mathbf{00}} \sum_{\mathbf{G'}\neq\mathbf{0}} B^{-1}_{\mathbf{GG'}} \tilde{\varepsilon}_{\mathbf{G'0}} \quad \text{(A15)}\\ \tilde{\varepsilon}^{-1}_{\mathbf{0G}} &= -\tilde{\varepsilon}^{-1}_{\mathbf{00}} \sum_{\mathbf{G'}\neq\mathbf{0}} \tilde{\varepsilon}_{\mathbf{0G'}} B^{-1}_{\mathbf{G'G}} \quad \text{(A16)}\\ \tilde{\varepsilon}^{-1}_{\mathbf{GG'}} &= B^{-1}_{\mathbf{GG'}} + \tilde{\varepsilon}^{-1}_{\mathbf{00}} \left(\sum_{\mathbf{G''}\neq\mathbf{0}} B^{-1}_{\mathbf{GG''}} \tilde{\varepsilon}_{\mathbf{G''0}}\right)\left(\sum_{\mathbf{G''}\neq\mathbf{0}} \tilde{\varepsilon}_{\mathbf{0G''}} B^{-1}_{\mathbf{G''G'}}\right) \quad \text{(A17)}\end{aligned}$$

Introducing the vectors $\mathbf{a}_{\mathbf{G}}$, $\mathbf{b}_{\mathbf{G}}$ and the tensor A given by

$$\begin{aligned}\mathbf{a}_{\mathbf{G}} &= -\sum_{\mathbf{G'}\neq\mathbf{0}} B^{-1}_{\mathbf{GG'}}\sqrt{v^{\mathrm{2D}}_{\mathbf{G'}}(\mathbf{0})}\mathbf{p}_{\mathbf{G'}} \quad \text{(A18)}\\ \mathbf{b}_{\mathbf{G}} &= -\sum_{\mathbf{G'}\neq\mathbf{0}} \sqrt{v^{\mathrm{2D}}_{\mathbf{G'}}(\mathbf{0})}\mathbf{s}_{\mathbf{G'}} B^{-1}_{\mathbf{G'G}} \quad \text{(A19)}\\ \mathsf{A} &= -\mathsf{P} - \sum_{\mathbf{G}\neq\mathbf{0}} \sqrt{v^{\mathrm{2D}}_{\mathbf{G}}(\mathbf{q}_\parallel)}\mathbf{s}_{\mathbf{G}} \otimes \mathbf{a}_{\mathbf{G}}, \quad \text{(A20)}\end{aligned}$$

where $\otimes$ denotes the tensor product, the long wavelength limit of the inverse dielectric function is seen to be given by

$$\begin{aligned}\tilde{\varepsilon}^{-1}_{\mathbf{00}}(\mathbf{q}_\parallel \to \mathbf{0}) &= \frac{1}{1 + 4\pi(1 - e^{-|\mathbf{q}_\parallel| L/2})\hat{\mathbf{q}}_\parallel \cdot \mathsf{A}\hat{\mathbf{q}}_\parallel} \quad \text{(A21)}\\ \tilde{\varepsilon}^{-1}_{\mathbf{G0}}(\mathbf{q}_\parallel \to \mathbf{0}) &= -\frac{\sqrt{4\pi(1 - e^{-|\mathbf{q}_\parallel| L/2})}\hat{\mathbf{q}}_\parallel \cdot \mathbf{a}_{\mathbf{G}}}{1 + 4\pi(1 - e^{-|\mathbf{q}_\parallel| L/2})\hat{\mathbf{q}}_\parallel \cdot \mathsf{A}\hat{\mathbf{q}}_\parallel} \quad \text{(A22)}\\ \tilde{\varepsilon}^{-1}_{\mathbf{0G}}(\mathbf{q}_\parallel \to \mathbf{0}) &= -\frac{\sqrt{4\pi(1 - e^{-|\mathbf{q}_\parallel| L/2})}\hat{\mathbf{q}}_\parallel \cdot \mathbf{b}_{\mathbf{G}}}{1 + 4\pi(1 - e^{-|\mathbf{q}_\parallel| L/2})\hat{\mathbf{q}}_\parallel \cdot \mathsf{A}\hat{\mathbf{q}}_\parallel} \quad \text{(A23)}\\ \tilde{\varepsilon}^{-1}_{\mathbf{GG'}}(\mathbf{q}_\parallel \to \mathbf{0}) &= B^{-1}_{\mathbf{GG'}} \\ &\quad + \frac{4\pi(1 - e^{-|\mathbf{q}_\parallel| L/2})(\hat{\mathbf{q}}_\parallel \cdot \mathbf{a}_{\mathbf{G}})(\hat{\mathbf{q}}_\parallel \cdot \mathbf{b}_{\mathbf{G'}})}{1 + 4\pi(1 - e^{-|\mathbf{q}_\parallel| L/2})\hat{\mathbf{q}}_\parallel \cdot \mathsf{A}\hat{\mathbf{q}}_\parallel} \quad \text{(A24)}\end{aligned}$$

Inserting these expressions in the equation for the correlation part of the screened potential, Eq. (2), we see that the head and wings are given by

$$\begin{aligned}
\overline{W}_{\mathbf{00}}(\mathbf{q}_\parallel \to \mathbf{0}) &= v_{\mathbf{0}}^{\mathrm{2D}}(\mathbf{q}_\parallel)\left[\tilde{\varepsilon}^{-1}_{\mathbf{00}}(\mathbf{q}_\parallel) - 1\right] \\
&= -\left(\frac{4\pi(1-e^{-|\mathbf{q}_\parallel|L/2})}{|\mathbf{q}_\parallel|}\right)^2 \frac{\hat{\mathbf{q}}_\parallel \cdot \mathsf{A}\hat{\mathbf{q}}_\parallel}{1+4\pi(1-e^{-|\mathbf{q}_\parallel|L/2})\hat{\mathbf{q}}_\parallel \cdot \mathsf{A}\hat{\mathbf{q}}_\parallel} \qquad \text{(A25)} \\
\overline{W}_{\mathbf{G0}}(\mathbf{q}_\parallel \to \mathbf{0}) &= \sqrt{v_{\mathbf{G}}^{\mathrm{2D}}(\mathbf{0})}\tilde{\varepsilon}^{-1}_{\mathbf{G0}}(\mathbf{q}_\parallel)\sqrt{v_{\mathbf{0}}^{\mathrm{2D}}(\mathbf{q}_\parallel)} \\
&= -\frac{4\pi(1-e^{-|\mathbf{q}_\parallel|L/2})}{|\mathbf{q}_\parallel|}\frac{\sqrt{v_{\mathbf{G}}^{\mathrm{2D}}(\mathbf{0})}\hat{\mathbf{q}}_\parallel \cdot \mathbf{a}_{\mathbf{G}}}{1+4\pi(1-e^{-|\mathbf{q}_\parallel|L/2})\hat{\mathbf{q}}_\parallel \cdot \mathsf{A}\hat{\mathbf{q}}_\parallel} \qquad \text{(A26)} \\
\overline{W}_{\mathbf{0G}}(\mathbf{q}_\parallel \to \mathbf{0}) &= \sqrt{v_{\mathbf{G}}^{\mathrm{2D}}(\mathbf{0})}\tilde{\varepsilon}^{-1}_{\mathbf{G0}}(\mathbf{q}_\parallel)\sqrt{v_{\mathbf{0}}^{\mathrm{2D}}(\mathbf{q}_\parallel)} \\
&= -\frac{4\pi(1-e^{-|\mathbf{q}_\parallel|L/2})}{|\mathbf{q}_\parallel|}\frac{\sqrt{v_{\mathbf{G}}^{\mathrm{2D}}(\mathbf{0})}\hat{\mathbf{q}}_\parallel \cdot \mathbf{b}_{\mathbf{G}}}{1+4\pi(1-e^{-|\mathbf{q}_\parallel|L/2})\hat{\mathbf{q}}_\parallel \cdot \mathsf{A}\hat{\mathbf{q}}_\parallel}. \qquad \text{(A27)}
\end{aligned}$$

and the body also gets a correction and becomes

$$\begin{aligned}
\overline{W}_{\mathbf{GG'}}(\mathbf{q}_\parallel \to \mathbf{0}) &= \sqrt{v_{\mathbf{G}}^{\mathrm{2D}}(\mathbf{0})v_{\mathbf{G'}}^{\mathrm{2D}}(\mathbf{0})}\left[\varepsilon^{-1}_{\mathbf{GG'}}(\mathbf{q}_\parallel) - \delta_{\mathbf{GG'}}\right] \\
&= \sqrt{v_{\mathbf{G}}^{\mathrm{2D}}(\mathbf{0})v_{\mathbf{G'}}^{\mathrm{2D}}(\mathbf{0})}\left[B^{-1}_{\mathbf{GG'}} - \delta_{\mathbf{GG'}} + \frac{4\pi(1-e^{-|\mathbf{q}_\parallel|L/2})(\hat{\mathbf{q}}_\parallel \cdot \mathbf{a}_{\mathbf{G}})(\hat{\mathbf{q}}_\parallel \cdot \mathbf{b}_{\mathbf{G'}})}{1+4\pi(1-e^{-|\mathbf{q}_\parallel|L/2})\hat{\mathbf{q}}_\parallel \cdot \mathsf{A}\hat{\mathbf{q}}_\parallel}\right]. \qquad \text{(A28)}
\end{aligned}$$

The $\mathbf{q}_\parallel = \mathbf{0}$ value is then $\overline{W}_{\mathbf{00}}(\mathbf{q}_\parallel = \mathbf{0}) = -(2\pi L)^2 \hat{\mathbf{q}}_\parallel \cdot \mathsf{A}\hat{\mathbf{q}}_\parallel$.

Introducing the dimensionless quantity $\mathbf{x} = \mathbf{q}_\parallel L/2$ and the rotational average $A = \frac{1}{2\pi}\int_0^{2\pi} \hat{\mathbf{x}}(\phi) \cdot \mathsf{A}\hat{\mathbf{x}}(\phi) d\phi$ we can define a new dimensionless function, $\tilde{w}(\mathbf{x})$

$$\tilde{w}(\mathbf{x}) = -\left(\frac{1-e^{-|\mathbf{x}|}}{|\mathbf{x}|}\right)^2 \frac{1}{1+4\pi A(1-e^{-|\mathbf{x}|})}. \qquad \text{(A29)}$$

from which the head of $\overline{W}$ can be written $\overline{W}(2\mathbf{x}/L) = (2\pi L)^2 A\tilde{w}(\mathbf{x})$. Using the rotational average of A is an approximate way of taking anisotropy of the system into account. It is possible to go beyond the approximation and to do it exactly, as was done for the three-dimensional case in reference 32 and 52, however this goes beyond the purpose of the present paper.

It is evident that the polar integral of Eq. (A29), over a small circle with radius r_{Ω_0}, cannot be evaluated analytically:

$$\begin{aligned}
&\int_{\Omega_0} \tilde{w}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \\
&-2\pi\int_0^{r_{\Omega_0}} \left(\frac{1-e^{-\mathbf{x}}}{\mathbf{x}}\right)^2 \frac{1}{1+4\pi A(1-e^{-\mathbf{x}})}\mathbf{x}\,\mathrm{d}\mathbf{x}. \qquad \text{(A30)}
\end{aligned}$$

It is however noticed that the function $\tilde{y}(\mathbf{x}) = \frac{1}{1+(1+4\pi A)\mathbf{x}}$ agrees very well with the integrand for small $\mathbf{x}$. It has the same first order Taylor expansion and it is integrable. This yields

$$\begin{aligned}
&\int_{\Omega_0} \tilde{w}(\mathbf{x})\,\mathrm{d}\mathbf{x} \approx -2\pi\int_0^{r_{\Omega_0}} \frac{\mathbf{x}}{1+(1+4\pi A)\mathbf{x}}\,\mathrm{d}\mathbf{x} = \\
&\frac{-2\pi(4\pi A r_{\Omega_0} + r_{\Omega_0} - \ln(4\pi A r_{\Omega_0} + r_{\Omega_0} + 1)}{(4\pi A+1)^2}. \qquad \text{(A31)}
\end{aligned}$$

Since the expression in Eq. (A31) only holds for small $\mathbf{x}$, it is generally not valid in the entire Ω_0 region.
The rotational average A is also used in the expressions for the wings and body of $\overline{W}$ but the rest of the terms are not treated analytically in the same way as the head but sampled numerically within the Ω_0 region.

[1] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, Phys. Rev. Lett. **105**, 136805 (2010).

[2] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.-Y.

Chim, G. Galli, and F. Wang, Nano Lett. **10**, 1271 (2010).
[3] H. Zeng, G.-B. Liu, J. Dai, Y. Yan, B. Zhu, R. He, L. Xie, S. Xu, X. Chen, W. Yao, and X. Cui, Sci Rep **3** (2013), 10.1038/srep01608.
[4] C. Zhang, A. Johnson, C.-L. Hsu, L.-J. Li, and C.-K. Shih, Nano Lett. **14**, 2443 (2014).
[5] Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, and M. S. Strano, Nature Nanotechnology **7**, 699 (2012).
[6] L. Britnell, R. M. Ribeiro, A. Eckmann, R. Jalil, B. D. Belle, A. Mishchenko, Y.-J. Kim, R. V. Gorbachev, T. Georgiou, S. V. Morozov, A. N. Grigorenko, A. K. Geim, C. Casiraghi, A. H. C. Neto, and K. S. Novoselov, Science **340**, 1311 (2013).
[7] M. Bernardi, M. Palummo, and J. C. Grossman, Nano Lett. **13**, 3664 (2013).
[8] W. G. Aulbur, L. Jönsson, and J. W. Wilkins, in *Solid State Physics*, Vol. Volume 54, edited by Henry Ehrenreich and Frans Spaepen (Academic Press, 1999) pp. 1–218.
[9] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).
[10] L. Hedin, Phys. Rev. **139**, A796 (1965).
[11] M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. **55**, 1418 (1985).
[12] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. Lett. **56**, 2415 (1986).
[13] M. Shishkin and G. Kresse, Phys. Rev. B **74**, 035101 (2006).
[14] T. Kotani and M. van Schilfgaarde, Solid State Communications **121**, 461 (2002).
[15] A. Marini, C. Hogan, M. Grüning, and D. Varsano, Computer Physics Communications **180**, 1392 (2009).
[16] C. Rostgaard, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B **81**, 085103 (2010).
[17] F. Caruso, P. Rinke, X. Ren, A. Rubio, and M. Scheffler, Phys. Rev. B **88**, 075105 (2013).
[18] F. Bruneval, Phys. Rev. Lett. **103**, 176403 (2009).
[19] X. Blase, C. Attaccalite, and V. Olevano, Phys. Rev. B **83**, 115103 (2011).
[20] S. Ismail-Beigi, Phys. Rev. B **73**, 233103 (2006).
[21] F. Huser, T. Olsen, and K. S. Thygesen, Phys. Rev. B **88**, 245309 (2013).
[22] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, Phys. Rev. Lett. **115**, 119901 (E) (2015).
[23] F. A. Rasmussen and K. S. Thygesen, J. Phys. Chem. C **119**, 13169 (2015).
[24] L. Keldysh, Jetp Lett. **29**, 658 (1979), wOS:A1979JN00600015.
[25] P. Cudazzo, I. V. Tokatly, and A. Rubio, Phys. Rev. B **84**, 085406 (2011).
[26] R. Sundararaman and T. A. Arias, Phys. Rev. B **87**, 165122 (2013).
[27] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
[28] R. M. Pick, M. H. Cohen, and R. M. Martin, Phys. Rev. B **1**, 910 (1970).
[29] J. Deslippe, G. Samsonidze, D. A. Strubbe, M. Jain, M. L. Cohen, and S. G. Louie, Computer Physics Communications **183**, 1269 (2012).
[30] C. Freysoldt, P. Eggert, P. Rinke, A. Schindlmayr, and M. Scheffler, Phys. Rev. B **77**, 235428 (2008).
[31] C. A. Rozzi, D. Varsano, A. Marini, E. K. U. Gross, and A. Rubio, Phys. Rev. B **73**, 205119 (2006).
[32] C. Freysoldt, P. Eggert, P. Rinke, A. Schindlmayr, R. W. Godby, and M. Scheffler, Computer Physics Communications **176**, 1 (2007).
[33] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, Phys. Rev. B **71**, 035109 (2005).
[34] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, and K. W. Jacobsen, Journal of Physics: Condensed Matter **22**, 253202 (2010).
[35] F. Hüser, T. Olsen, and K. S. Thygesen, Phys. Rev. B **87**, 235132 (2013).
[36] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[37] M. L. Tiago, S. Ismail-Beigi, and S. G. Louie, Phys. Rev. B **69**, 125212 (2004).
[38] B.-C. Shih, Y. Xue, P. Zhang, M. L. Cohen, and S. G. Louie, Phys. Rev. Lett. **105**, 146401 (2010).
[39] J. Klimeš, M. Kaltak, and G. Kresse, Phys. Rev. B **90**, 075125 (2014).
[40] J. Yan, J. J. Mortensen, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B **83**, 245122 (2011).
[41] Z. Y. Zhu, Y. C. Cheng, and U. Schwingenschlögl, Phys. Rev. B **84**, 153402 (2011).
[42] A. Ramasubramaniam, Phys. Rev. B **86**, 115409 (2012).
[43] T. Cheiwchanchamnangij and W. R. L. Lambrecht, Phys. Rev. B **85**, 205302 (2012).
[44] H.-P. Komsa and A. V. Krasheninnikov, Phys. Rev. B **86**, 241201 (2012).
[45] A. Molina-Sánchez, D. Sangalli, K. Hummer, A. Marini, and L. Wirtz, Phys. Rev. B **88**, 045412 (2013).
[46] H. Shi, H. Pan, Y.-W. Zhang, and B. I. Yakobson, Phys. Rev. B **87**, 155304 (2013).
[47] H. J. Conley, B. Wang, J. I. Ziegler, R. F. Haglund, S. T. Pantelides, and K. I. Bolotin, Nano Lett. **13**, 3626 (2013).
[48] A. R. Klots, A. K. M. Newaz, B. Wang, D. Prasai, H. Krzyzanowska, J. Lin, D. Caudel, N. J. Ghimire, J. Yan, B. L. Ivanov, K. A. Velizhanin, A. Burger, D. G. Mandrus, N. H. Tolk, S. T. Pantelides, and K. I. Bolotin, Sci. Rep. **4** (2014), 10.1038/srep06608.
[49] N. Berseneva, A. Gulans, A. V. Krasheninnikov, and R. M. Nieminen, Phys. Rev. B **87**, 035404 (2013).
[50] V. Tran, R. Soklaski, Y. Liang, and L. Yang, Phys. Rev. B **89**, 235319 (2014).
[51] A. Baldereschi and E. Tosatti, Phys. Rev. B **17**, 4710 (1978).
[52] C. Friedrich, S. Blügel, and A. Schindlmayr, Phys. Rev. B **81**, 125102 (2010).

Paper IV

Beyond the GW method: Simple vertex correction improves band energies of bulk and two-dimensional materials

Per S. Schmidt, Christopher E. Patrick and Kristian S. Thygesen

Submitted.

Beyond the GW method: Simple vertex correction improves band energies of bulk and two-dimensional materials

Per S. Schmidt, Christopher E. Patrick,* and Kristian S. Thygesen
Center for Atomic-scale Materials Design (CAMD),
Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark
(Dated: July 28, 2017)

The GW self-energy method has long been recognized as the gold standard for quasiparticle (QP) calculations of solids in spite of the fact that the neglect of vertex corrections and use of a DFT starting point lacks rigorous justification. In this work we remedy this situation by including a simple vertex correction that is consistent with an LDA starting point. We analyse the effect of the self-energy by splitting it into short-range and long-range terms which are shown to govern respectively the center and size of the band gap. The vertex mainly improves the short-range correlations and therefore has a small effect on the band gap, while it shifts the band gap *center* up in energy by around 0.5 eV in good agreement with experiments. Our analysis also explains how the different relative importance of short- and long-range interactions in structures of different dimensionality is reflected in their QP energies. Inclusion of the vertex comes at practically no extra computational cost and even improves the basis set convergence compared to GW. The QP-rALDA method thus provides an efficient and rigorous improvement over the GW approximation and sets a new standard for quasiparticle calculations in solids.

The GW approximation [1–4], introduced by Hedin in 1965 [5] remains the most widely used method for quasiparticle (QP) calculations of semiconductors and insulators. Over the years it has been extensively applied to inorganic solids [6–8] and more recently also to molecules [9–12] and atomically thin two-dimensional (2D) materials [13–15].

The GW self-energy can be obtained by iterating Hedin's equations once starting from $\Sigma = 0$ (i.e. the Hartree approximation). This produces the trivial vertex function $\Gamma(1,2,3) = \delta(1,2)\delta(1,3)$, which corresponds to invoking the time-dependent Hartree approximation for the dynamical screening (i.e. the random phase approximation (RPA)). For this approach to be consistent, the Green's function which should be used for the calculation of the self-energy is the Hartree G. This is known to be a poor approximation, and instead practical GW calculations follow a "best G, best W" philosophy [3]. Most often one uses a non-interacting G_0 from density functional theory (DFT) and evaluates W within the RPA from the polarisability $\chi_0 = G_0G_0$. This approximation is referred to as G_0W_0 and has shown to yield reasonably good, although somewhat underestimated, band gaps [6, 16]. Carrying out self-consistency in the Green's function only, GW_0, has been found to improve the band gaps [17]. Iterating to full self-consistency in both the Green's function and screened interaction, GW, systematically overestimates the band gaps and worsens the agreement with experiments [17].

The most obvious way to go beyond the GW approximation is to perform another iteration of Hedin's equations starting from $\Sigma = iGW$. Neglecting derivatives of W this produces the kernel $\delta\Sigma(1,2)/\delta G(3,4) = iW(1,2,3,4)$, which is known from the Bethe-Salpeter Equation. The four-point nature of this kernel makes it difficult to invert the vertex equation, $\Gamma = \delta + KGG\Gamma$, without loss of accuracy. Instead one can perform a single iteration of the vertex equation to obtain $\Gamma = \delta + WGG$, which leads to a self-energy consisting of a second-order screened exchange term in addition to the usual iGW term. Gruneis *et al.* have shown, using a static approximation for W in the vertex, that this $GW\Gamma^1$ approximation, performed in a fully self-consistent manner, leads to significant improvements for band gaps and ionization potentials of solids [18]. From a theoretical point of view this is a highly satisfactory result. The drawback is the higher complexity of the formalism and the concomitant loss of physical transparency as well as the significant computational overhead as compared to the GW method.

Time-dependent density-functional theory (TDDFT) in principle offers a framework for including exchange-correlation (xc)-effects in the dynamical response via a two-point vertex function rather than the computationally challenging three-point vertex function that arises naturally in the diagrammatic many-body formalism. While it appears attractive to use TDDFT derived vertex functions for many-body calculations, progress along these lines has been hindered by the poor quality of the local xc-kernels derived from standard local xc-potentials. However, recent work has shown that a simple renormalization of the adiabatic LDA xc-kernel can overcome these problems and yield a dramatic improvement over the RPA for total energy calculations based on the adiabatic connection fluctuation dissipation theorem (ACDFT) [19–21].

Here we show that the renormalized adiabatic LDA (rALDA) kernel, when introduced in Hedin's equations, produces a simple two-point vertex function that leads to systematically improved QP energies for a range of semi-

conductors and insulators. The most striking effect of the vertex is that it raises the absolute QP energies from G_0W_0 by around 0.5 eV while the gaps are almost unaffected. These effects can be traced to an improved description of the short range correlation hole and thus the (absolute) correlation energy of electrons in the ground state.

As originally observed by Hybertsen and Louie [3], it is possible to start the iterative solution of Hedin's equation not with $\Sigma = 0$ (which leads to the GW approximation), but rather with a local xc-potential: $\Sigma^0(1,2) = \delta(1,2)v_{xc}(1)$. As shown by Del Sole *et al.* [22] this leads to a self-energy of the form

$$\Sigma(1,2) = iG(1,2)\widetilde{W}(1,2), \tag{1}$$

where

$$\widetilde{W} = v[1 - \chi^0(v + f_{xc})]^{-1} \tag{2}$$

and $f_{xc}(1,2) = \delta v_{xc}(1)/\delta n(2)$ is the adiabatic xc-kernel. Crucially, $\widetilde{W}(1,2)$ is the screened *effective* potential at 2 generated by a charge at 1. It consists of the bare potential plus the induced Hartree and xc-potential. It is thus the potential felt by a (Kohn-Sham) electron in the system. For comparison the potential felt by a classical test charge is the bare potential screened only by the induced Hartree potential:

$$\widehat{W} = v + v[1 - \chi^0(v + f_{xc})]^{-1}\chi^0 v \tag{3}$$

Using $\widehat{W}$ in Eq. (1) corresponds to including the vertex in the polarisability P (or irreducible response function) but neglecting it in the self-energy. We shall refer to the use of $\widetilde{W}$ or $\widehat{W}$ in Eq. (1) as $G_0W_0\Gamma$ and G_0W_0P, respectively. The subscripts indicate that the self-energies are evaluated non-self-consistently starting from DFT. Note that in contrast to the GW approximation, which strictly should be based on the Hartree G, the use of a DFT starting point is perfectly justified within the $G_0W_0\Gamma$ theory.

Here, the exchange part of the rALDA kernel [19–21] is employed in Eq.(2) and (3). The rALDA kernel is defined for the HEG by setting $f_{xc}^{\mathrm{rALDA}}[n](q) = f_{xc}^{\mathrm{ALDA}}[n]$ for $q < 2k_F[n]$ and $-v(q)$ otherwise. This results in a non-local kernel with the (almost) exact asymptotic $q \to \infty$ behaviour and without the divergences of the ALDA kernel [21]. The rALDA kernel improves the RPA description of bond energies in molecules and solids and more importantly for the present work is that the rALDA kernel provides a dramatic improvement of absolute correlation energies compared to RPA. For example, the RPA correlation energy of the HEG is 0.3-0.5 eV/electron too negative while the rALDA error is below 0.03 eV/electron. Similar trends are seen for small atoms, molecules [20] and bulk silicon [21].

We have calculated the QP band gaps, ionization potentials (IP) and electron affinities (EA) for a range of semiconductors and insulators using five different approximations to the self-energy: (i) conventional G_0W_0 (ii) eigenvalue self-consistent GW_0 (iii) full eigenvalue self-consistent GW (iv) $G_0W_0\Gamma$ and (v) G_0W_0P. The non-self-consistent calculations employed an LDA starting point and the exchange only rALDA kernel was employed. The QP calculations for the bulk and 2D crystals were performed in their experimental geometries using the GPAW code [23]. Band edge positions with respect to vacuum were determined by aligning the Hartree potential at the nuclei in the bulk calculations to that inside a slab with surface orientation and reconstruction as reported in available experimental studies. The considered surfaces are (111) 2×1 for Si in the diamond structure, (100) for MgO and LiF in the rocksalt structure and (110) for the rest of the compounds in the zinc-blende structure.

	LDA	G_0W_0	GW_0	$G_0W_0P_0$	$G_0W_0\Gamma_0$	Exp.
MgO	4.68	7.70	8.16	7.10	7.96	7.83
CdS	0.86	1.76	2.27	1.84	1.87	2.42
LiF	8.83	14.00	14.75	13.25	14.21	14.20
SiC	1.31	2.54	2.72	2.38	2.57	2.40
Si	0.52	1.23	1.34	1.16	1.29	1.17
C	4.10	5.74	5.97	5.62	5.69	5.81
BN	4.36	6.54	6.81	6.27	6.60	6.66
AlP	1.44	2.48	2.67	2.34	2.51	2.45
ML-MoS_2	1.71	2.47	2.61	2.28	2.47	2.50
ML-$MoSe_2$	1.43	2.08	2.23	1.99	2.07	2.31
ML-WS_2	1.33	2.75	3.07	2.56	2.81	2.72
MAE	1.81	0.15	0.24	0.34	0.14	-
MSE	-1.81	-0.11	0.19	-0.34	-0.04	-

TABLE I: Band gaps obtained using different self-energy approximations (see text). Experimental values for C and BN were corrected for zero-point motion (0.33 eV, 0.41 eV respectively [24]). For the experimental gaps of the 2D structures, see Section 6 in the Supplemental Material [25]. Spin-orbit corrections have been added to the 2D structures.

Table I shows the band gaps obtained with the different methods and their deviations from experimental reference values. In agreement with previous findings G_0W_0 underestimates the experimental band gaps slightly while GW_0 generally overestimates. The overestimation becomes even larger in GW (see Supplemental Material [25]) which is therefore not considered further in this work. $G_0W_0\Gamma$ shows the best agreement with experiments, and with a mean signed error of only -0.04 eV the small deviations are not systematic. Including the vertex only in the polarisability (G_0W_0P) leads to significant underestimation of the gaps.

In Fig. 1 we show the absolute positions of the valence band maximum (VBM) and conduction band minimum (CBM) relative to vacuum. The most striking effect of

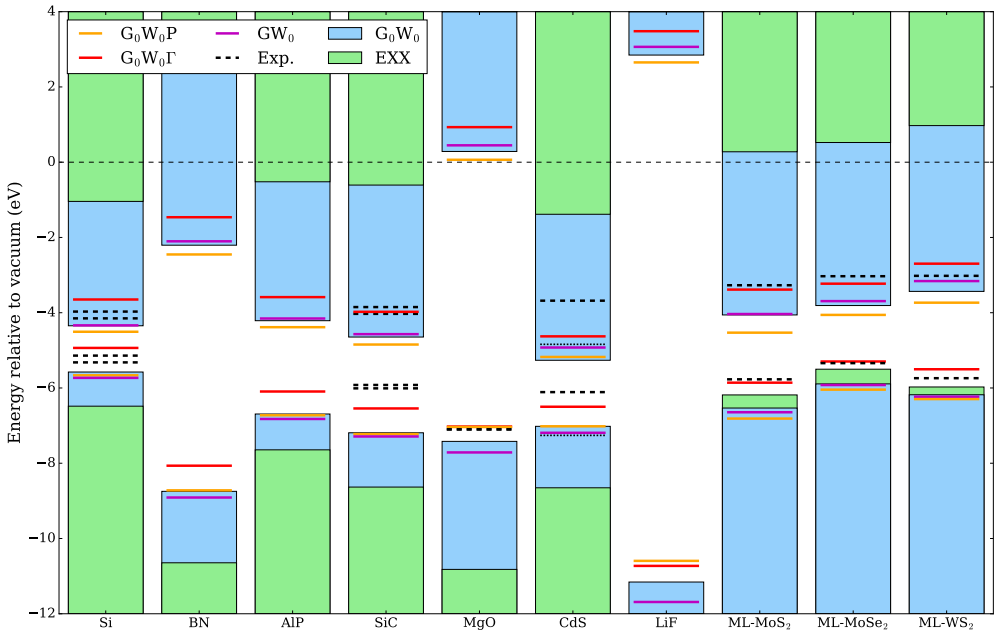


FIG. 1: IP and EA of a range of 3D and 2D semiconductors calculated with EXX (green), G_0W_0 (blue), $G_0W_0\Gamma$ (red), G_0W_0P (orange), GW_0 (magenta) and compared with experimental values where available (black) [25, 26].

including the vertex is a systematic upshift of the band edges by around 0.5 eV. Remarkably, this upshift leads to a better overall agreement with experiments (dashed black lines). The upshift of band energies is not observed when the vertex is included only in the polarisability, i.e. when employing a test charge-test charge screened interaction (G_0W_0P). Moreover, no systematic upshift of the band edges is observed for the self-consistent GW flavours which also employ test charge-test charge screening. We conclude that the upshift of band energies originates from the presence of the vertex in the self-energy, i.e. the use of a test charge-electron screened interaction.

In the following we analyse our results from a total energy perspective focusing on the G_0W_0 and $G_0W_0\Gamma$ methods. From Koopmans' theorem it follows that the highest occupied and lowest unoccupied QP energies can be expressed as

$$\varepsilon_N^{\mathrm{QP}} = \varepsilon_N^{\mathrm{HF}} + E_c[N] - E_c[N-1] \quad (4)$$

$$\varepsilon_{N+1}^{\mathrm{QP}} = \varepsilon_{N+1}^{\mathrm{HF}} + E_c[N+1] - E_c[N] \quad (5)$$

where $\varepsilon^{\mathrm{HF}}$ are the Hartree-Fock single particle energies (evaluated on Kohn-Sham orbitals) and $E_c[N]$ is the correlation energy of the N-particle ground state. The latter can be calculated from the ACDFT, which can be cast in the form

$$E_c = -\int_0^1 d\lambda \int_0^\infty \frac{d\omega}{2\pi} \mathrm{Tr}\{\chi^0(i\omega)(\widetilde{W}^\lambda(i\omega) - v)\} \quad (6)$$

Here $\widetilde{W}^\lambda$ equals the screened test charge-electron interaction of Eq. (2). Setting $f_{xc} = 0$ we have $\widetilde{W} = W$ and E_c becomes the RPA correlation energy. Assuming no orbital relaxations (which is justified for an extended periodic crystal), Niquet *et al.* [27] have shown that the ionization potential (IP) and electron affinity (EA) calculated as total energy differences with the ACDFT-RPA equal the highest occupied and lowest unoccupied QP energies from G_0W_0, respectively (when setting the renormalization factor Z to unity). In the same way it can be shown, at least for an exchange only kernel, that the IP and EA obtained from the ACDFT with $\widetilde{W}$ from Eq. (2), equal the respective QP band edges obtained from $G_0W_0\Gamma$ when Γ is the vertex corresponding to f_x (see the Supplemental Material [25] for a proof). These results represent a generalization of Koopmans' theorem of Hartree-Fock theory.

In general, HF is known to significantly overestimate the band gap of solids (see Fig. 1). Comparing with Eqs. (4-5) this means that the correlation energy in the $N \pm 1$ states must be larger (more negative) than the correlation energy in the neutral N-particle ground state. It might

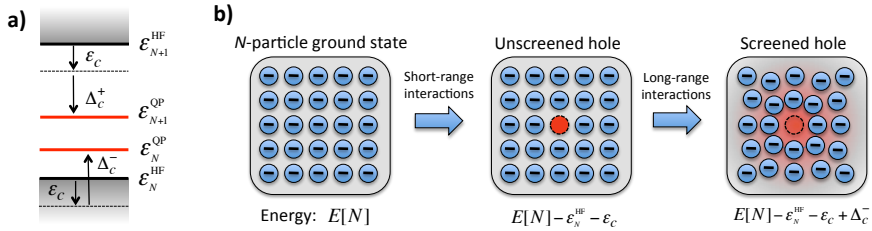


FIG. 2: (a) Schematic illustration of the different contributions to the highest occupied and lowest unoccupied QP levels of a semiconductor. (b) The energy cost of removing a valence electron consists of the Hartree-Fock energy ($\varepsilon_N^{\rm HF}$), the correlation energy of an electron in the ground state (ε_c), and a stabilising screening contribution ($\Delta_c^{\pm}$). The two latter are predominantly of short-range and long-range nature, respectively.

seem surprising that $E_c[N-1] < E_c[N]$ since naively one would expect the correlation energy to be a monotonic decreasing function of N. However, the addition of an electron/hole to the system changes its character from insulating to metallic and this entails an increase in the correlation energy. To make this idea more explicit we can split the change in correlation energy into two terms: the correlation energy per electron in the neutral ground state ($\varepsilon_c \equiv E_c[N]/N < 0$) and a remainder representing the extra correlation energy due to the insulator-metal transition ($\Delta_c^{+/-} \equiv E_c[N \pm 1] - (E_c[N] \pm \varepsilon_c)$). With these definitions we can write

$$\varepsilon_N^{\rm QP} = \varepsilon_N^{\rm HF} + \varepsilon_c - \Delta_c^- \tag{7}$$

$$\varepsilon_{N+1}^{\rm QP} = \varepsilon_{N+1}^{\rm HF} + \varepsilon_c + \Delta_c^+ \tag{8}$$

The relations are illustrated in Fig. 2(a). Clearly, the effect of ε_c is to downshift the band edges from their HF positions while the $\Delta_c^{\pm}$ closes the gap. In the quasiparticle picture, $\Delta_c^{\pm}$ represent the screening of the added electron/hole, see Fig. 2(b), and we shall therefore refer to them as screening terms. By its stabilization of the final states (the $N \pm 1$ states) the effect of the screening terms is similar to that of orbital relaxations in finite systems, yet the underlying physics is completely different: orbital relaxations are vanishingly small in periodic crystals and occur even in non-correlated theories like HF. In contrast $\Delta_c^{\pm}$ describes a pure correlation effect and does not vanish in infinite, periodic systems.

We find it useful to analyse the QP energies in terms of the band gap and the band gap center. These are related to ε_c and $\Delta_c^{\pm}$ by

$$E_{\rm gap}^{\rm QP} = E_{\rm gap}^{\rm HF} + (\Delta_c^- + \Delta_c^+) \tag{9}$$

$$E_{\rm cen}^{\rm QP} = E_{\rm cen}^{\rm HF} + \varepsilon_c + (\Delta_c^+ - \Delta_c^-)/2 \tag{10}$$

The correlation contribution to the gap is determined only by the screening terms $\Delta_c^{\pm}$. From the close agreement between the G_0W_0 and $G_0W_0\Gamma$ band gaps (red columns in Fig. 3) we conclude that the vertex has little effect on the (sum of the) screening terms. In contrast,

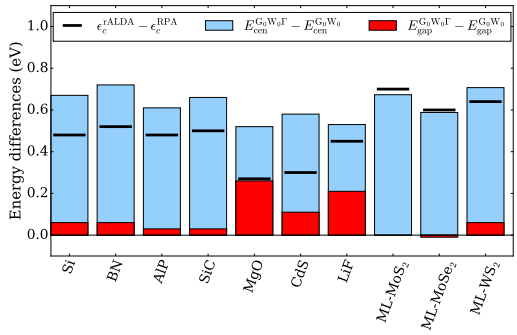


FIG. 3: The difference in band gap, band gap center, and ϵ_c upon inclusion of the rALDA kernel.

the band gap center also depends on the ground state correlation energy, ε_c. We have calculated ε_c for all the investigated materials using the RPA and rALDA total energy methods (see the Supplemental Material [25] for details). In Fig. 3 we compare the difference between the RPA and rALDA calculated ε_c (black line) with the difference between the G_0W_0 and $G_0W_0\Gamma$ calculated band gap centers (blue columns). The rather close agreement shows that the main difference in band gap center can be ascribed to ε_c. It is now clear that the upshift of QP energies obtained with $G_0W_0\Gamma$ originates from the smaller (less negative) correlation energy of electrons in the neutral ground state predicted by rALDA compared to RPA. The well documented superiority of the rALDA over the RPA for the description of ground state correlation energies, in combination with the improved agreement with experimental band energies (Fig. 1) constitutes strong evidence that our QP-rALDA scheme represents a genuine improvement over the GW approximation.

We have seen that the dominant effect of the vertex correction is to shift the band gap center while the band

gap itself is less affected. Physically, the main effect of the rALDA kernel is to modify the effective Coulomb interaction at short distances. More precisely, given a density variation, δn, the corresponding induced electron potential, $\delta v_{\mathrm{Hxc}} = (v + f_{xc})\delta n$, is generally weaker than the bare Hartree potential $\delta v_H = v\delta n$, because $v(q) + f_{xc}(q) < v(q)$. However, by definition of the rALDA kernel, the reduction is stronger for larger q, which translates to shorter distances in real space. From these observations we can conclude that the QP band gap is mainly determined by long-range interactions while the band gap center is sensitive to the short-range interactions. This agrees well with the quasiparticle picture illustrated in Fig. 2: Namely, adding a particle/hole without accounting for the screening represents a local (short-range) perturbation while the screening of the added charge is a long-range process. While the rALDA kernel mainly reduces the short-range interactions it also reduces the long-range components slightly. This leads to a slightly weaker long-range screening (smaller Δ_c) and slightly larger band gaps as seen in Table I.

Returning to Fig. 1 we note that for the 2D materials Hartree-Fock predicts a lower IP than the GW methods in clear contrast to the situation for bulk solids. This anomalous behaviour is a result of the relatively more important effect of short- compared to long-range correlations in reduced dimensions. Indeed, the dielectric function of a 2D semiconductor approaches unity in the long wavelength limit, which reduces the screening terms $\Delta_c^{\pm}$. At the same time we find that the 2D materials present the largest values for ε_c of all the materials (see Table 4 in the Supplemental Material [25]).

Finally, we mention that the reduction of large q components by the rALDA kernel improves convergence with respect to plane waves and number of unoccupied states (see Fig. 1 in the Supplemental Material [25]).

In conclusion, we have demonstrated that a more accurate description of short-range correlations in QP calculations can be obtained with a simple TDDFT-inspired vertex function. Inclusion of the vertex improves the agreement with experimental data for the absolute band energies of bulk and two-dimensional semiconductors. Moreover, it justifies the use of DFT as a starting point for non-self-consistent QP calculations and is thus formally more rigorous than the G_0W_0@DFT approach. Importantly, these advantages come without increasing the numerical complexity or computational cost compared to G_0W_0 calculations.

The Center for Nanostructured Graphene (CNG) is sponsored by the Danish Research Foundation, Project DNRF103.

* Present address: Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom

[1] W. G. Aulbur, L. Jönsson, and J. W. Wilkins, in *Solid State Physics*, Vol. Volume 54, edited by Henry Ehrenreich and Frans Spaepen (Academic Press, 1999) pp. 1–218.
[2] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).
[3] M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. **55**, 1418 (1985).
[4] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. Lett. **56**, 2415 (1986).
[5] L. Hedin, Phys. Rev. **139**, A796 (1965).
[6] M. Shishkin and G. Kresse, Phys. Rev. B **74**, 035101 (2006).
[7] T. Kotani and M. van Schilfgaarde, Solid State Communications **121**, 461 (2002).
[8] A. Marini, C. Hogan, M. Grüning, and D. Varsano, Computer Physics Communications **180**, 1392 (2009).
[9] C. Rostgaard, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B **81**, 085103 (2010).
[10] F. Caruso, P. Rinke, X. Ren, A. Rubio, and M. Scheffler, Phys. Rev. B **88**, 075105 (2013).
[11] F. Bruneval, Phys. Rev. Lett. **103**, 176403 (2009).
[12] X. Blase, C. Attaccalite, and V. Olevano, Phys. Rev. B **83**, 115103 (2011).
[13] F. Hüser, T. Olsen, and K. S. Thygesen, Phys. Rev. B **87**, 235132 (2013).
[14] F. A. Rasmussen, P. S. Schmidt, K. T. Winther, and K. S. Thygesen, Phys. Rev. B **94**, 155406 (2016).
[15] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, Phys. Rev. Lett. **115**, 119901 (2015).
[16] P. Scherpelz, M. Govoni, I. Hamada, and G. Galli, Journal of Chemical Theory and Computation **12**, 3523 (2016), pMID: 27331614.
[17] M. Shishkin and G. Kresse, Phys. Rev. B **75**, 235102 (2007).
[18] A. Grüneis, G. Kresse, Y. Hinuma, and F. Oba, Phys. Rev. Lett. **112**, 096401 (2014).
[19] T. Olsen and K. S. Thygesen, Phys. Rev. B **86**, 081103 (2012).
[20] T. Olsen and K. S. Thygesen, Phys. Rev. B **88**, 115131 (2013).
[21] C. E. Patrick and K. S. Thygesen, The Journal of Chemical Physics **143**, 102802 (2015).
[22] R. Del Sole, L. Reining, and R. W. Godby, Phys. Rev. B **49**, 8024 (1994).
[23] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, and K. W. Jacobsen, Journal of Physics: Condensed Matter **22**, 253202 (2010).
[24] S. Poncé, Y. Gillet, J. L. Janssen, A. Marini, M. Verstraete, and X. Gonze, The Journal of Chemical Physics **143**, 102813 (2015).
[25] See Supplemental Material at [URL]..

[26] W. Mönch, *Semiconductor Surfaces and Interfaces* (Springer, 2001).

[27] Y. M. Niquet and X. Gonze, Phys. Rev. B **70**, 245115 (2004).

Supplementary information

Beyond the GW method: Simple vertex correction improves band energies of bulk and two-dimensional materials

Per S. Schmidt, Christopher E. Patrick and Kristian S. Thygesen

1 From rALDA total energies to GWΓ

Using the adiabatic connection and fluctuation-dissipation theorem (ACFDT), the exact correlation energy of the system can be written in terms of the interacting response function ($\chi^\lambda(i\omega)$) of a system with a scaled Coulomb interaction, $v \to \lambda v$:

$$E_c = -\int_0^1 d\lambda \int_0^\infty \frac{d\omega}{2\pi} \mathrm{Tr}\Big(v \Big[\chi^\lambda(i\omega) - \chi^{\mathrm{KS}}(i\omega) \Big] \Big), \tag{1}$$

where $\chi^{\mathrm{KS}}(i\omega)$ is the response function of the non-interacting Kohn-Sham system. χ^λ can in principle be obtained from the Dyson equation

$$\chi^\lambda(i\omega) = \chi^{\mathrm{KS}}(i\omega) + \chi^{\mathrm{KS}}(i\omega) \Big[\lambda v + f_{xc}^\lambda(i\omega) \Big] \chi^\lambda(i\omega),$$

where all the complicated correlation effects has been transferred into $f_{xc}^\lambda(i\omega)$, which needs to be approximated.
It can be shown, that any pure exchange kernel must have the property $f_x^\lambda[n](\mathbf{r}, \mathbf{r}', i\omega) = \lambda f_x[n](\mathbf{r}, \mathbf{r}', i\omega)$ making it possible to carry out the λ-integration analytically:

$$E_c = \int_0^\infty \frac{d\omega}{2\pi} \mathrm{Tr}\Big[v f_{Hx}^{-1}(i\omega) \ln\Big[1 - \chi^{\mathrm{KS}}(i\omega) f_{Hx}(i\omega) \Big] + v\chi^{\mathrm{KS}}(i\omega) \Big], \tag{2}$$

where $f_{Hx}(i\omega) = v + f_x(i\omega)$.
Denoting the change in $\chi^{\mathrm{KS}}(i\omega)$ when adding one electron to the lowest unoccupied KS orbital by $\delta\chi^{\mathrm{KS}}(i\omega)$,

$$\delta\chi^{\mathrm{KS}}(\mathbf{r}, \mathbf{r}', i\omega) = \phi_c^*(\mathbf{r}) G_0(\mathbf{r}, \mathbf{r}', \epsilon_c + i\omega) \phi_c(\mathbf{r}') + \text{c.c.},$$

assuming that the density doesn't change by the addition of one electron $\Rightarrow$ $f_{Hx}(i\omega)$ doesn't change, we calculate the correlation contribution to the electron

affinity (AE):

$$
\begin{aligned}
AE_c &= E_c[N] - E_c[N+1] \\
&= \int_0^\infty \frac{d\omega}{2\pi} \mathrm{Tr}\Big[v f_{Hx}^{-1}(i\omega) \ln\Big[1 - \chi^{\mathrm{KS}}(i\omega) f_{Hx}(i\omega)\Big] + v\chi^{\mathrm{KS}}(i\omega) \\
&- \Big[v f_{Hx}^{-1}(i\omega) \ln\Big[1 - \Big(\chi^{\mathrm{KS}}(i\omega) + \delta\chi^{\mathrm{KS}}(i\omega)\Big) f_{Hx}(i\omega)\Big] + v\Big(\chi^{\mathrm{KS}}(i\omega) + \delta\chi^{\mathrm{KS}}(i\omega)\Big)\Big]\Big] \\
&= -\int_0^\infty \frac{d\omega}{2\pi} \mathrm{Tr}\Big[v f_{Hx}^{-1}(i\omega) \Big(\ln\Big[1 - \chi^{\mathrm{KS}}(i\omega) f_{Hx}(i\omega) - \delta\chi^{\mathrm{KS}}(i\omega) f_{Hx}(i\omega)\Big] \\
&- \ln\Big[1 - \chi^{\mathrm{KS}}(i\omega) f_{Hx}(i\omega)\Big]\Big) + v\delta\chi^{\mathrm{KS}}(i\omega)\Big] \\
&= -\int_0^\infty \frac{d\omega}{2\pi} \mathrm{Tr}\Big[v f_{Hx}^{-1}(i\omega) \ln\Big[1 - \delta\chi^{\mathrm{KS}}(i\omega) f_{Hx}(i\omega)\Big(1 - \chi^{\mathrm{KS}}(i\omega) f_{Hx}(i\omega)\Big)^{-1}\Big] + v\delta\chi^{\mathrm{KS}}(i\omega)\Big]
\end{aligned}
$$

Adding and subtracting $v\Big[1 - \chi^{\mathrm{KS}}(i\omega) f_{Hx}(i\omega)\Big]^{-1} \delta\chi^{\mathrm{KS}}(i\omega)$, we can split AE_c in two terms, $AE_c = AE_c^{\mathrm{QP}} + AE_c'$ and following the arguments of Niquet et al.[1] AE_c' vanishes, leaving us with

$$
\begin{aligned}
AE_c &= \int_0^\infty \frac{d\omega}{2\pi} \mathrm{Tr}\Big[\Big(v\Big[1 - \chi^{\mathrm{KS}}(i\omega) f_{Hx}(i\omega)\Big]^{-1} - v\Big)\delta\chi^{\mathrm{KS}}(i\omega)\Big] \\
&= \int_0^\infty \frac{d\omega}{2\pi} \mathrm{Tr}\Big[\Big(W(i\omega) - v\Big)\delta\chi^{\mathrm{KS}}(i\omega)\Big]
\end{aligned}
$$

which is exactly the correlation contribution to the conduction band matrix element of the GW self-energy in the GWΓ method:

$$AE_c = \langle\phi_c|\Sigma_c(\epsilon_c)|\phi_c\rangle$$

where

$$\Sigma_c(\mathbf{r}, \mathbf{r}', \epsilon_c) = -\int \frac{d\omega}{2\pi} G_0(\mathbf{r}, \mathbf{r}', \epsilon_c + i\omega) W(\mathbf{r}, \mathbf{r}', i\omega)$$

and $W(i\omega) = v\Big[1 - \chi^{\mathrm{KS}}(i\omega) f_{Hx}(i\omega)\Big]^{-1}$.

2 Structures and computational details

	Structure	Lattice constant (Å)	k-points
MgO	rocksalt	4.212	$8 \times 8 \times 8$
SiC	zincblende	4.350	$8 \times 8 \times 8$
LiF	rocksalt	4.024	$8 \times 8 \times 8$
CdS	zincblende	5.818	$8 \times 8 \times 8$
Si	diamond	5.431	$8 \times 8 \times 8$
C	diamond	3.567	$8 \times 8 \times 8$
BN	zincblende	3.615	$8 \times 8 \times 8$
AlP	zincblende	5.451	$8 \times 8 \times 8$

Table 1: Bulk structures.

	Structure	Lattice constant (Å)	Layer thickness (Å)	k-points
MoS_2	2H	3.160	3.170	$18 \times 18 \times 1$
$MoSe_2$	2H	3.289	3.340	$18 \times 18 \times 1$
WS_2	2H	3.153	3.360	$18 \times 18 \times 1$

Table 2: 2D structures.

The quasiparticle and total correlation energy calculations were carried out in the experimental structures reported in Table 1 and 2 employing a $8 \times 8 \times 8$ ($18 \times 18 \times 1$) **k**-point grid for the bulk (2D) materials and extrapolating to infinite plane wave cutoff energy and number of bands included in construction of χ_0. For the quasiparticle calculations of the 2D materials we employed a recently developed method for treating the critical $\mathbf{q} \to \mathbf{0}$ limit of the screened interaction while avoiding spurious interactions with neighbouring supercells [2]. The response functions and screened interactions were evaluated along the real frequency axis using a non-linear frequency grid.
For the total energy calculations the frequency integration was evaluated using 16 frequency points and a Gaussian quadrature. The details of the implementation is described in [3]. The kernel used in the rALDA calculations were constructed using the wavevector average as described in [4].
For the 2D structures, 15 Å of vacuum was used in the out-of-plane direction between repeated unit cells.
ϵ_c is in turn defined as the total correlation energy per number of *valence* electrons.

3 $G_0W_0\Gamma$ method: Improved convergence

In addition to improved bandgaps and absolute band positions, we report an improved convergence when including the rALDA kernel in the $G_0W_0\Gamma$ method. The kernel suppresses higher plane waves thus making the $G_0W_0\Gamma$ method converge faster with respect to plane wave basis size and the number of unoccupied bands, as shown in Fig. 1 for the case of bulk BN. On the y-axis we show the difference in the band gap compared to that from a calculation at a cutoff of 50 eV with the corresponding number of bands. The improved convergence also manifests itself in the QP corrections to the individual bands.

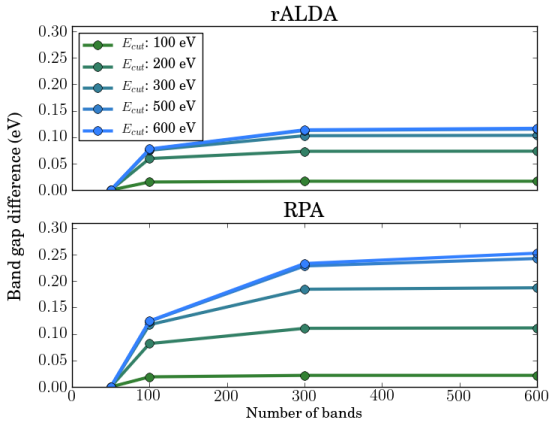


Figure 1: Convergence of the band gap in BN with respect to plane wave cutoff and the number of bands included using the RPA (bottom) and rALDA (top) kernel.

4 Eigenvalue self-consistency

	GW_0	GW_0P_0	$GW_0\Gamma_0$	GW	GWP_0	$GW\Gamma_0$	Exp.
MgO	8.16	7.52	8.21	9.21	8.56	9.47	7.83
CdS	2.27	2.03	2.34	2.73	2.44	2.83	2.42
LiF	14.75	14.02	14.90	16.29	15.59	16.46	14.20
SiC	2.72	2.52	2.73	3.06	2.80	3.11	2.40
Si	1.34	1.23	1.40	1.48	1.38	1.60	1.17
C	5.97	5.74	5.88	6.38	6.16	6.33	5.81
BN	6.81	6.50	6.84	7.44	7.08	7.49	6.66
AlP	2.67	2.48	2.67	3.00	2.74	3.01	2.45
MAE	0.26	0.16	0.27	0.83	0.48	0.92	-
MSE	0.22	-0.11	0.25	0.83	0.48	0.92	-

Table 3: Bandgaps calculated with eigenvalue self-consistency in G (GW_0) and in both G and W (GW).

5 ϵ_c and $\Delta_c^{\pm}$

The absolute values of ϵ_c and $\Delta_c^{\pm}$ as well as the sum and differences of $\Delta_c^{\pm}$ contributing to the band gaps and centers respectively.

	ϵ_c from total energies			Δ_c^+		Δ_c^-		$\Delta_c^- + \Delta_c^+$			$(\Delta_c^+ - \Delta_c^-)/2$		
	RPA	rALDA	Δ	G_0W_0	$G_0W_0\Gamma_0$	G_0W_0	$G_0W_0\Gamma_0$	G_0W_0	$G_0W_0\Gamma_0$	Δ	G_0W_0	$G_0W_0\Gamma_0$	Δ
Si	-1.53	-1.05	0.48	-1.77	-1.56	-2.44	-2.60	-4.21	-4.16	0.05	0.34	0.52	0.19
BN	-1.75	-1.23	0.52	-3.02	-2.79	-3.65	-3.81	-6.67	-6.60	0.07	0.32	0.51	0.20
AlP	-1.50	-1.02	0.48	-2.19	-2.05	-2.45	-2.57	-4.64	-4.62	0.02	0.13	0.26	0.13
SiC	-1.64	-1.14	0.50	-2.40	-2.23	-3.08	-3.23	-5.48	-5.46	0.02	0.34	0.50	0.16
MgO	-1.00	-0.73	0.27	-2.84	-2.46	-4.40	-4.53	-7.24	-6.99	0.25	0.78	1.04	0.26
CdS	-1.13	-0.83	0.30	-2.75	-2.41	-2.76	-2.98	-5.51	-5.39	0.12	0.01	0.29	0.28
LiF	-1.59	-1.14	0.45	-1.87	-1.69	-5.96	-5.94	-7.83	-7.63	0.2	2.05	2.13	0.08
ML-MoS_2	-1.97	-1.27	0.70	-2.37	-2.39	-1.62	-1.60	-3.99	-3.99	0.00	-0.38	-0.40	0.02
ML-$MoSe_2$	-1.87	-1.27	0.60	-2.36	-2.48	-1.58	-1.47	-3.94	-3.95	0.01	-0.39	-0.51	0.12
ML-WS_2	-1.73	-1.09	0.64	-2.44	-2.40	-1.76	-1.74	-4.20	-4.14	0.06	-0.34	-0.33	0.01

Table 4

6 2D materials: IP, EA and gaps

VBM, CBM and band gaps for three 2D semiconductors calculated with various methods, relative to vacuum. The HSE, G_0W_0 and $G_0W_0\Gamma$ are all non-self-

consistent calculations on top of LDA orbitals and eigenvalues. The experimental VBM are from currently unpublished work by A. D. Mohite, T. Ohta *et al.* to appear in ACS Nano. The band gap of $MoSe_2$ is from [5] where they have measured a gap of 2.18 ± 0.04 eV for $MoSe_2$ on top of bilayer graphene. The effect of the substrate is calculated to be a lowering of the band gap of 0.13 eV, giving a band gap of 2.31 eV for free-standing $MoSe_2$.
The band gap of 2.5 eV for free-standing MoS_2 is from [6]. In [7] they report a band gap of 2.18 ± 0.05 eV for MoS_2 on top of quartz. Comparing the two numbers, quartz is expected to lower the gap by 0.32 eV. In [7] the band gap of WS_2 on top of quartz is reported to be 2.40 ± 0.06 eV. Assuming the same substrate effect, the band gap of free-standing WS_2 must be 2.72 eV.
The numbers printed are including spin-orbit corrections. These are a splitting of the VB by 0.15, 0.19, 0.45 eV and of the CB by 0.00, 0.02, 0.02 eV for MoS_2, $MoSe_2$ and WS_2 respectively [8].

	LDA	HSE	G_0W_0	G_0W_0P	$G_0W_0\Gamma$	GW_0	Exp.
				MoS_2			
VBM	-6.07	-6.19	-6.53	-6.81	-5.86	-6.65	-5.77
CBM	-4.35	-3.93	-4.06	-4.53	-3.39	-4.04	-3.27
Gap	1.71	2.25	2.47	2.28	2.47	2.61	2.50
				$MoSe_2$			
VBM	-5.48	-5.58	-5.89	-6.04	-5.30	-5.92	-5.34
CBM	-4.06	-3.64	-3.81	-4.06	-3.23	-3.69	-3.03
Gap	1.43	1.94	2.08	1.99	2.07	2.23	2.31
				WS_2			
VBM	-5.59	-5.78	-6.18	-6.29	-5.50	-6.23	-5.74
CBM	-4.26	-3.95	-3.43	-3.73	-2.70	-3.16	-3.02
Gap	1.33	1.83	2.75	2.56	2.81	3.07	2.72

References

[1] Y. M. Niquet and X. Gonze, "Band-gap energy in the random-phase approximation to density-functional theory," *Phys. Rev. B*, vol. 70, p. 245115, Dec 2004.

[2] F. A. Rasmussen, P. S. Schmidt, K. T. Winther, and K. S. Thygesen, "Efficient many-body calculations for two-dimensional materials using exact limits for the screened potential: Band gaps of $\mathbf{mos}_2$, h-bn, and phosphorene," *Phys. Rev. B*, vol. 94, p. 155406, Oct 2016.

[3] T. Olsen and K. S. Thygesen, "Random phase approximation applied to solids, molecules, and graphene-metal interfaces: From van der waals to covalent bonding," *Phys. Rev. B*, vol. 87, p. 075111, Feb 2013.

[4] C. E. Patrick and K. S. Thygesen, "Adiabatic-connection fluctuation-dissipation dft for the structural properties of solids—the renormalized alda and electron gas kernels," *The Journal of Chemical Physics*, vol. 143, no. 10, 2015.

[5] M. M. Ugeda, A. J. Bradley, S.-F. Shi, F. H. da Jornada, Y. Zhang, D. Y. Qiu, W. Ruan, S.-K. Mo, Z. Hussain, Z.-X. Shen, F. Wang, S. G. Louie, and M. F. Crommie, "Giant bandgap renormalization and excitonic effects in a monolayer transition metal dichalcogenide semiconductor," *Nature Materials*, vol. 13, p. 1091–1095, Aug 2014.

[6] A. R. Klots, A. K. M. Newaz, B. Wang, D. Prasai, H. Krzyzanowska, J. Lin, D. Caudel, N. J. Ghimire, J. Yan, B. L. Ivanov, K. A. Velizhanin, A. Burger, D. G. Mandrus, N. H. Tolk, S. T. Pantelides, and K. I. Bolotin, "Probing excitonic states in suspended two-dimensional semiconductors by photocurrent spectroscopy," *Scientific Reports*, vol. 4, p. 6608, Oct 2014.

[7] H. M. Hill, A. F. Rigosi, K. T. Rim, G. W. Flynn, and T. F. Heinz, "Band alignment in mos2/ws2 transition metal dichalcogenide heterostructures probed by scanning tunneling microscopy and spectroscopy," *Nano Letters*, vol. 16, no. 8, pp. 4831–4837, 2016. PMID: 27298270.

[8] F. A. Rasmussen and K. S. Thygesen, "Computational 2d materials database: Electronic structure of transition-metal dichalcogenides and oxides," *The Journal of Physical Chemistry C*, vol. 119, no. 23, pp. 13169–13183, 2015.

APPENDIX A

Appendix

A.1 Norm-conserving setups for W and Mo

Performance:		
	Norm-conserving	
	Mo d	W d
Fitness factor	8.8	5.9
Log. deriv.	0.3	0.5
fcc - a_0	0.00	0.00
fcc - 10%	0.01	0.01
fcc - 20%	0.01	0.01
rocksalt - a_0	0.00	0.00
rocksalt - 10%	0.04	0.01
rocksalt - 20%	0.04	0.02
Iterations	35	31
Cutoff	614	483
Eggbox	0.0026	0.0012
Radii	0.0	0.0
Parameters:		
Radii: s	1.95	2.20
Radii: p	2.30	2.10
Radii: d	1.95	2.50
Energies: s	-	-
Energies: p	-	-
Energies: d	-0.3	0.5
Radii: Local pot.	1.80	1.75

Table A.1: Performance and parameters for the GA generated norm-conserving Mo and W setups used in this work. For Mo the 4s, 5s, 4p, 5p, 4d are treated as valence electrons and for W it is the 5s, 6s, 5p, 6p, 5d. To both setups, an extra d partial wave is added.

Bibliography

[1] J. Wood, "The top ten advances in materials science," *Materials Today*, vol. 11, no. 1, pp. 40 – 45, 2008.

[2] "This month in physics history: April 25, 1954: Bell labs demonstrates the first practical silicon solar cell," *APS News*, vol. 18, April 2009.

[3] W. Shockley and H. J. Queisser, "Detailed balance limit of efficiency of p-n junction solar cells," *Journal of Applied Physics*, vol. 32, no. 3, pp. 510–519, 1961.

[4] F. Dimroth, T. N. D. Tibbits, M. Niemeyer, F. Predan, P. Beutel, C. Karcher, E. Oliva, G. Siefer, D. Lackner, P. Fuß-Kailuweit, A. W. Bett, R. Krause, C. Drazek, E. Guiot, J. Wasselin, A. Tauzin, and T. Signamarcheix, "Four-junction wafer-bonded concentrator solar cells," *IEEE Journal of Photovoltaics*, vol. 6, pp. 343–349, Jan 2016.

[5] W. Hafez and M. Feng, "Experimental demonstration of pseudomorphic heterojunction bipolar transistors with cutoff frequencies above 600ghz," *Applied Physics Letters*, vol. 86, no. 15, p. 152101, 2005.

[6] A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Challenges for Density Functional Theory," *Chem. Rev.*, Jan. 2012.

[7] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, "Electric field effect in atomically thin carbon films," *Science*, vol. 306, no. 5696, pp. 666–669, 2004.

[8] G. R. Bhimanapati, Z. Lin, V. Meunier, Y. Jung, J. Cha, S. Das, D. Xiao, Y. Son, M. S. Strano, V. R. Cooper, L. Liang, S. G. Louie, E. Ringe, W. Zhou, S. S. Kim, R. R. Naik, B. G. Sumpter, H. Terrones, F. Xia, Y. Wang, J. Zhu, D. Akinwande, N. Alem, J. A. Schuller, R. E. Schaak, M. Terrones, and J. A. Robinson, "Recent advances in two-dimensional

materials beyond graphene," *ACS Nano*, vol. 9, no. 12, pp. 11509–11539, 2015. PMID: 26544756.

[9] A. C. Ferrari, F. Bonaccorso, V. Fal'ko, K. S. Novoselov, S. Roche, P. Boggild, S. Borini, F. H. L. Koppens, V. Palermo, N. Pugno, J. A. Garrido, R. Sordan, A. Bianco, L. Ballerini, M. Prato, E. Lidorikis, J. Kivioja, C. Marinelli, T. Ryhanen, A. Morpurgo, J. N. Coleman, V. Nicolosi, L. Colombo, A. Fert, M. Garcia-Hernandez, A. Bachtold, G. F. Schneider, F. Guinea, C. Dekker, M. Barbone, Z. Sun, C. Galiotis, A. N. Grigorenko, G. Konstantatos, A. Kis, M. Katsnelson, L. Vandersypen, A. Loiseau, V. Morandi, D. Neumaier, E. Treossi, V. Pellegrini, M. Polini, A. Tredicucci, G. M. Williams, B. Hee Hong, J.-H. Ahn, J. Min Kim, H. Zirath, B. J. van Wees, H. van der Zant, L. Occhipinti, A. Di Matteo, I. A. Kinloch, T. Seyller, E. Quesnel, X. Feng, K. Teo, N. Rupesinghe, P. Hakonen, S. R. T. Neil, Q. Tannock, T. Lofwander, and J. Kinaret, "Science and technology roadmap for graphene, related two-dimensional crystals, and hybrid systems," *Nanoscale*, vol. 7, pp. 4598–4810, 2015.

[10] A. A. Balandin, S. Ghosh, W. Bao, I. Calizo, D. Teweldebrhan, F. Miao, and C. N. Lau, "Superior thermal conductivity of single-layer graphene," *Nano Letters*, vol. 8, no. 3, pp. 902–907, 2008. PMID: 18284217.

[11] A. S. Mayorov, R. V. Gorbachev, S. V. Morozov, L. Britnell, R. Jalil, L. A. Ponomarenko, P. Blake, K. S. Novoselov, K. Watanabe, T. Taniguchi, and A. K. Geim, "Micrometer-scale ballistic transport in encapsulated graphene at room temperature," *Nano Letters*, vol. 11, no. 6, pp. 2396–2399, 2011. PMID: 21574627.

[12] C. R. Dean, A. F. Young, I. Meric, C. Lee, L. Wang, S. Sorgenfrei, K. Watanabe, T. Taniguchi, P. Kim, K. L. Shepard, and J. Hone, "Boron nitride substrates for high-quality graphene electronics," *Nature Nanotechnology*, vol. 5, pp. 722–726, Oct. 2010.

[13] L. A. Ponomarenko, A. K. Geim, A. A. Zhukov, R. Jalil, S. V. Morozov, K. S. Novoselov, I. V. Grigorieva, E. H. Hill, V. V. Cheianov, V. I. Fal'Ko, K. Watanabe, T. Taniguchi, and R. V. Gorbachev, "Tunable metal-insulator transition in double-layer graphene heterostructures," *Nature Physics*, vol. 7, pp. 958–961, Dec. 2011.

[14] L. Britnell, R. V. Gorbachev, R. Jalil, B. D. Belle, F. Schedin, M. I. Katsnelson, L. Eaves, S. V. Morozov, A. S. Mayorov, N. M. R. Peres,

A. H. Castro Neto, J. Leist, A. K. Geim, L. A. Ponomarenko, and K. S. Novoselov, "Electron tunneling through ultrathin boron nitride crystalline barriers," *Nano Letters*, vol. 12, no. 3, pp. 1707–1710, 2012. PMID: 22380756.

[15] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, "Atomically thin mos_2: A new direct-gap semiconductor," *Phys. Rev. Lett.*, vol. 105, p. 136805, Sep 2010.

[16] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.-Y. Chim, G. Galli, and F. Wang, "Emerging photoluminescence in monolayer mos_2," *Nano Letters*, vol. 10, no. 4, pp. 1271–1275, 2010. PMID: 20229981.

[17] A. K. Geim and I. V. Grigorieva, "Van der Waals heterostructures," *ArXiv e-prints*, July 2013.

[18] P. Cudazzo, I. V. Tokatly, and A. Rubio, "Dielectric screening in two-dimensional insulators: Implications for excitonic and impurity states in graphane," *Phys. Rev. B*, vol. 84, p. 085406, Aug 2011.

[19] H. Kroemer, "A proposed class of hetero-junction injection lasers," *Proceedings of the IEEE*, vol. 51, pp. 1782–1783, Dec 1963.

[20] H. Liu, "Ammonia synthesis catalyst 100 years: Practice, enlightenment and challenge," *Chinese Journal of Catalysis*, vol. 35, no. 10, pp. 1619 – 1640, 2014.

[21] R. French and S. Czernik, "Catalytic pyrolysis of biomass for biofuels production," *Fuel Processing Technology*, vol. 91, no. 1, pp. 25 – 32, 2010.

[22] K. C. Taylor, *Automobile Catalytic Converters*, pp. 119–170. Berlin, Heidelberg: Springer Berlin Heidelberg, 1984.

[23] J. Wellendorff, T. L. Silbaugh, D. Garcia-Pintos, J. K. Nørskov, T. Bligaard, F. Studt, and C. T. Campbell, "A benchmark database for adsorption bond energies to transition metal surfaces and comparison to selected {DFT} functionals," *Surface Science*, vol. 640, pp. 36 – 44, 2015. Reactivity Concepts at Surfaces: Coupling Theory with Experiment.

[24] "Global gas flaring reduction partnership: Improving energy efficiency and mitigating impact on climate change," *The World Bank. Global Gas Flaring Reduction Partnership: Improving Energy Efficiency and Mitigating Impact on Climate Change*, 2011.

[25] P. Forster, V. Ramaswamy, P. Artaxo, T. Berntsen, R. Betts, D. Fahey, J. Haywood, J. Lean, D. Lowe, G. Myhre, J. Nganga, M. S. R. P. G. Raga, and R. V. Dorland *Changes in Atmospheric Constituents and in Radiative Forcing*, 2007.

[26] M. H. Ab Rahim, M. M. Forde, C. Hammond, R. L. Jenkins, N. Dimitratos, J. A. Lopez-Sanchez, A. F. Carley, S. H. Taylor, D. J. Willock, and G. J. Hutchings, "Systematic study of the oxidation of methane using supported gold palladium nanoparticles under mild aqueous conditions," *TOPICS IN CATALYSIS*, vol. 56, no. 18, 2013.

[27] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Phys. Rev.*, vol. 140, pp. A1133–A1138, Nov 1965.

[28] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, "Real-space grid implementation of the projector augmented wave method," *Phys. Rev. B*, vol. 71, p. 035109, Jan 2005.

[29] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, and K. W. Jacobsen, "Electronic structure calculations with gpaw: a real-space implementation of the projector augmented-wave method," *Journal of Physics: Condensed Matter*, vol. 22, no. 25, p. 253202, 2010.

[30] J. Kohanoff, *Electronic Structure Calculations for Solids and Molecules*. Cambridge University Press, 2006.

[31] V. Fock, "Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems," *Zeitschrift für Physik*, vol. 61, pp. 126–148, Jan 1930.

[32] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Phys. Rev.*, vol. 136, pp. B864–B871, Nov 1964.

[33] C. Fiolhais, F. Nogueira, and M. Marques, *A Primer in Density Functional Theory*. Springer, 2003.

[34] N. Argaman and G. Makov, "Density Functional Theory - an introduction," 1999.

[35] D. M. Ceperley and B. J. Alder, "Ground state of the electron gas by a stochastic method," *Phys. Rev. Lett.*, vol. 45, pp. 566–569, Aug 1980.

[36] J. Perdew, "Density functional theory and the band gap problem," *International Journal of Quantum Chemistry*, vol. 28, no. S19, pp. 497–523, 1985.

[37] M. Born and R. Oppenheimer, "Zur quantentheorie der molekeln," *Annalen der Physik*, vol. 389, no. 20, pp. 457–484, 1927.

[38] P. E. Blöchl, "Projector augmented-wave method," *Phys. Rev. B*, vol. 50, pp. 17953–17979, Dec 1994.

[39] D. Whitley and A. M. Sutton, *Genetic Algorithms — A Survey of Models and Methods*, pp. 637–671. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[40] C. Rostgaard, "Exact exchange in density functional calculations," Master's thesis, DTU, 2006.

[41] D. Vanderbilt, "Optimally smooth norm-conserving pseudopotentials," vol. 32, pp. 8412–8415, 01 1986.

[42] J. c. v. Klimeš, M. Kaltak, and G. Kresse, "Predictive *gw* calculations using plane waves and pseudopotentials," *Phys. Rev. B*, vol. 90, p. 075125, Aug 2014.

[43] K. Sarkar, M. Topsakal, N. A. W. Holzwarth, and R. M. Wentzcovitch, "Evolutionary optimization of PAW data-sets for accurate high pressure simulations," *Journal of Computational Physics*, vol. 347, pp. 39–55, Oct. 2017.

[44] D. Nabok, A. Gulans, and C. Draxl, "Accurate all-electron G_0W_0 quasiparticle energies employing the full-potential augmented plane-wave method," *Phys. Rev. B*, vol. 94, p. 035118, Jul 2016.

[45] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Phys. Rev. Lett.*, vol. 77, pp. 3865–3868, Oct 1996.

[46] J. Heyd, G. E. Scuseria, and M. Ernzerhof, "Erratum: "hybrid functionals based on a screened coulomb potential" [j. chem. phys. 118, 8207 (2003)]," *The Journal of Chemical Physics*, vol. 124, no. 21, p. 219906, 2006.

[47] F. Furche, "Molecular tests of the random phase approximation to the exchange-correlation energy functional," *Phys. Rev. B*, vol. 64, p. 195120, Oct 2001.

[48] M. Fuchs and X. Gonze, "Accurate density functionals: Approaches using the adiabatic-connection fluctuation-dissipation theorem," *Phys. Rev. B*, vol. 65, p. 235109, Jun 2002.

[49] X. Ren, P. Rinke, C. Joas, and M. Scheffler, "Random-phase approximation and its applications in computational chemistry and materials science," *Journal of Materials Science*, vol. 47, pp. 7447–7471, Nov 2012.

[50] A. Marini, P. García-González, and A. Rubio, "First-principles description of correlation effects in layered materials," *Phys. Rev. Lett.*, vol. 96, p. 136404, Apr 2006.

[51] X. Ren, P. Rinke, C. Joas, and M. Scheffler, "Random-phase approximation and its applications in computational chemistry and materials science," *J. Mat. Sci.*, 2012.

[52] T. Olsen and K. Thygesen, "Beyond the random phase approximation: Improved description of short range correlation by a renormalized adiabatic local density approximation," *Phys. Rev. B 88, 115131*, 2013.

[53] T. Olsen and K. S. Thygesen, "Accurate ground-state energies of solids and molecules from time-dependent density-functional theory," *Phys. Rev. Lett.*, vol. 112, p. 203001, May 2014.

[54] T. Olsen and K. S. Thygesen, "Beyond the random phase approximation: Improved description of short-range correlation by a renormalized adiabatic local density approximation," *Phys. Rev. B*, vol. 88, p. 115131, Sep 2013.

[55] T. Olsen and K. S. Thygesen, "Extending the random-phase approximation for electronic correlation energies: The renormalized adiabatic local density approximation," *Phys. Rev. B*, vol. 86, p. 081103, Aug 2012.

[56] C. E. Patrick and K. S. Thygesen, "Adiabatic-connection fluctuation-dissipation dft for the structural properties of solids—the renormalized alda and electron gas kernels," *The Journal of Chemical Physics*, vol. 143, no. 10, p. 102802, 2015.

[57] J. Harl, L. Schimka, and G. Kresse, "Assessing the quality of the random phase approximation for lattice constants and atomization energies of solids," *Phys. Rev. B*, vol. 81, Mar. 2010.

[58] T. Olsen and K. S. Thygesen, "Correlation energies from the Random Phase Approximation," *Lecture notes, Course 10305 - Advanced Solid State Physics*, 2013.

[59] J. Harl and G. Kresse, "Cohesive energy curves for noble gas solids calculated by adiabatic connection fluctuation-dissipation theory," *Phys. Rev. B*, vol. 77, Jan. 2008.

[60] J. Harl and G. Kresse, "Accurate bulk properties from approximate many-body techniques," *Phys. Rev. Lett.*, vol. 103, July 2009.

[61] H. Jiang and E. Engel, "Random-phase-approximation-based correlation energy functionals: Benchmark results for atoms," *Jour. Chem. Phys.*, vol. 127, Nov. 2007.

[62] J. P. Perdew and Y. Wang, "Pair-distribution function and its coupling-constant average for the spin-polarized electron gas," *Phys. Rev. B*, vol. 46, pp. 12947–12954, Nov 1992.

[63] L. Schimka, J. Harl, A. Stroppa, A. Grüneis, M. Marsman, F. Mittendorfer, and G. Kresse, "Accurate surface and adsorption energies from many-body perturbation theory," *Nature Materials*, vol. 9, no. 9, pp. 741–744, 2010.

[64] X. Ren, P. Rinke, and M. Scheffler, "Exploring the random phase approximation: Application to co adsorbed on cu(111)," *Phys. Rev. B*, vol. 80, p. 045402, Jul 2009.

[65] T. Olsen, J. Yan, J. J. Mortensen, and K. S. Thygesen, "Dispersive and covalent interactions between graphene and metal surfaces from the random phase approximation," *Phys. Rev. Lett.*, vol. 107, p. 156401, Oct 2011.

[66] M. Rohlfing and T. Bredow, "Binding energy of adsorbates on a noble-metal surface: Exchange and correlation effects," *Phys. Rev. Lett.*, vol. 101, p. 266106, Dec 2008.

[67] J. Ma, A. Michaelides, D. Alfè, L. Schimka, G. Kresse, and E. Wang, "Adsorption and diffusion of water on graphene from first principles," *Phys. Rev. B*, vol. 84, p. 033402, Jul 2011.

[68] H.-J. Kim, A. Tkatchenko, J.-H. Cho, and M. Scheffler, "Benzene adsorbed on si(001): The role of electron correlation and finite temperature," *Phys. Rev. B*, vol. 85, p. 041403, Jan 2012.

[69] T. Olsen and K. S. Thygesen, "Accurate ground-state energies of solids and molecules from time-dependent density-functional theory," *Phys. Rev. Lett.*, vol. 112, p. 203001, May 2014.

[70] L. Hedin and S. Lundqvist, "Effects of electron-electron and electron-phonon interactions on the one-electron states of solids," *Solid State Physics*, vol. 23, pp. 1 – 181, 1970.

[71] K. S. Singwi, M. P. Tosi, R. H. Land, and A. Sjölander, "Electron correlations at metallic densities," *Phys. Rev.*, vol. 176, pp. 589–599, Dec 1968.

[72] "Global gas flaring reduction partnership: Improving energy efficiency and mitigating impact on climate change," *The World Bank. Global Gas Flaring Reduction Partnership: Improving Energy Efficiency and Mitigating Impact on Climate Change*, 2011.

[73] P. Forster, V. Ramaswamy, P. Artaxo, T. Berntsen, R. Betts, D. Fahey, J. Haywood, J. Lean, D. Lowe, G. Myhre, J. Nganga, M. S. R. P. G. Raga, and R. V. Dorland *Changes in Atmospheric Constituents and in Radiative Forcing*, 2007.

[74] M. H. Ab Rahim, M. M. Forde, C. Hammond, R. L. Jenkins, N. Dimitratos, J. A. Lopez-Sanchez, A. F. Carley, S. H. Taylor, D. J. Willock, and G. J. Hutchings, "Systematic study of the oxidation of methane using supported gold palladium nanoparticles under mild aqueous conditions," *TOPICS IN CATALYSIS*, vol. 56, no. 18, 2013.

[75] R. Palkovits, M. Antonietti, P. Kuhn, A. Thomas, and F. Schüth, "Solid catalysts for the selective low-temperature oxidation of methane

to methanol," *Angewandte Chemie International Edition*, vol. 48, no. 37, pp. 6909–6912, 2009.

[76] J. Rossmeisl, Z.-W. Qu, H. Zhu, G.-J. Kroes, and J. Nørskov, "Electrolysis of water on oxide surfaces," *Journal of Electroanalytical Chemistry*, vol. 607, p. 83, 2007.

[77] J. K. Nørskov, J. Rossmeisl, A. Logadottir, L. Lindqvist, J. R. Kitchin, T. Bligaard, and H. Jónsson, "Origin of the overpotential for oxygen reduction at a fuel-cell cathode," *The Journal of Physical Chemistry B*, vol. 108, no. 46, pp. 17886–17892, 2004.

[78] M. Naguib, J. Come, B. Dyatkin, V. Presser, P.-L. Taberna, P. Simon, M. W. Barsoum, and Y. Gogotsi, "Mxene: a promising transition metal carbide anode for lithium-ion batteries," *Electrochemistry Communications*, vol. 16, no. 1, pp. 61 – 64, 2012.

[79] M. Naguib, M. Kurtoglu, V. Presser, J. Lu, J. Niu, M. Heon, L. Hultman, Y. Gogotsi, and M. W. Barsoum, "Two-dimensional nanocrystals produced by exfoliation of ti3alc2," *Advanced Materials*, vol. 23, no. 37, pp. 4248–4253, 2011.

[80] M. Pandey and K. S. Thygesen, "Two-dimensional mxenes as catalysts for electrochemical hydrogen evolution: A computational screening study," *The Journal of Physical Chemistry C*, vol. 121, no. 25, pp. 13593–13598, 2017.

[81] A. Latimer, H. Aljama, A. Kakekhani, J. S. Yoo, A. Kulkarni, C. Tsai, M. Garcia-Melchor, F. Abild-Pedersen, and J. K. Nørskov, "Mechanistic insights into heterogeneous methane activation," *Physical Chemistry Chemical Physics*, vol. 19, pp. 3575–3581, 2017.

[82] F. Aryasetiawan and O. Gunnarsson, "The GW method," *Reports on Progress in Physics*, vol. 61, pp. 237–312, Mar. 1998.

[83] W. Aulbur, L. Jonsson, and J. Wilkins, "Quasiparticle Caculations in Solids," *Lecture notes*, 2013.

[84] F. Hüser, T. Olsen, and K. S. Thygesen, "Quasiparticle gw calculations for solids, molecules, and two-dimensional materials," *Phys. Rev. B*, vol. 87, p. 235132, Jun 2013.

[85] P. Fulde, *Semiconductors and Insulators*, pp. 189–221. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995.

[86] L. Hedin, "New method for calculating the one-particle green's function with application to the electron-gas problem," *Phys. Rev.*, vol. 139, pp. A796–A823, Aug 1965.

[87] L. Hedin and S. Lundqvist, *Effects of Electron-Electron and Electron-Phonon Interactions on the One-Electron States of Solids*, vol. 23, pp. 1–181. United States: Academic Press, 1969. doi:10.1016/S0081-1947(08)60615-3.

[88] M. S. Hybertsen and S. G. Louie, "First-principles theory of quasiparticles: Calculation of band gaps in semiconductors and insulators," *Phys. Rev. Lett.*, vol. 55, pp. 1418–1421, Sep 1985.

[89] M. Shishkin and G. Kresse, "Implementation and performance of the frequency-dependent *gw* method within the paw framework," *Phys. Rev. B*, vol. 74, p. 035101, Jul 2006.

[90] P. Scherpelz, M. Govoni, I. Hamada, and G. Galli, "Implementation and validation of fully relativistic gw calculations: Spin–orbit coupling in molecules, nanocrystals, and solids," *Journal of Chemical Theory and Computation*, vol. 12, no. 8, pp. 3523–3544, 2016. PMID: 27331614.

[91] R. Sundararaman and T. A. Arias, "Regularization of the coulomb singularity in exact exchange by wigner-seitz truncated interactions: Towards chemical accuracy in nontrivial systems," *Phys. Rev. B*, vol. 87, p. 165122, Apr 2013.

[92] C. Friedrich, M. Betzinger, M. Schlipf, S. Blügel, and A. Schindlmayr, "Hybrid functionals and g w approximation in the flapw method," *Journal of Physics: Condensed Matter*, vol. 24, no. 29, p. 293201, 2012.

[93] W. Gao, W. Xia, X. Gao, and P. Zhang, "Speeding up gw calculations to meet the challenge of large scale quasiparticle predictions," *Scientific Reports*, vol. 6, no. 36849.

[94] M. Govoni and G. Galli, "Large scale gw calculations," *Journal of Chemical Theory and Computation*, vol. 11, no. 6, pp. 2680–2696, 2015. PMID: 26575564.

[95] M. Shishkin and G. Kresse, "Self-consistent *gw* calculations for semiconductors and insulators," *Phys. Rev. B*, vol. 75, p. 235102, Jun 2007.

[96] S. Ismail-Beigi, "Truncation of periodic image interactions for confined systems," *Phys. Rev. B*, vol. 73, p. 233103, Jun 2006.

[97] C. A. Rozzi, D. Varsano, A. Marini, E. K. U. Gross, and A. Rubio, "Exact coulomb cutoff technique for supercell calculations," *Phys. Rev. B*, vol. 73, p. 205119, May 2006.

[98] A. Grüneis, G. Kresse, Y. Hinuma, and F. Oba, "Ionization potentials of solids: The importance of vertex corrections," *Phys. Rev. Lett.*, vol. 112, p. 096401, Mar 2014.

[99] R. Del Sole, L. Reining, and R. W. Godby, "Gw," *Phys. Rev. B*, vol. 49, pp. 8024–8028, Mar 1994.

[100] M. Shishkin, M. Marsman, and G. Kresse, "Accurate quasiparticle spectra from self-consistent gw calculations with vertex corrections," *Phys. Rev. Lett.*, vol. 99, p. 246403, Dec 2007.

[101] Y. M. Niquet and X. Gonze, "Band-gap energy in the random-phase approximation to density-functional theory," *Phys. Rev. B*, vol. 70, p. 245115, Dec 2004.

[102] L. J. Sham and M. Schlüter, "Density-functional theory of the energy gap," *Phys. Rev. Lett.*, vol. 51, pp. 1888–1891, Nov 1983.

[103] Y. Niquet and M. Fuchs.

[104] J. Klimeš and G. Kresse, "Kohn-sham band gaps and potentials of solids from the optimised effective potential method within the random phase approximation," *The Journal of Chemical Physics*, vol. 140, no. 5, p. 054516, 2014.

[105] M. V. Ganduglia-Pirovano, A. Hofmann, and J. Sauer, "Oxygen vacancies in transition metal and rare earth oxides: Current state of understanding and remaining challenges," *Surface Science Reports*, vol. 62, no. 6, pp. 219 – 270, 2007.

[106] C. Di Valentin, G. Pacchioni, and A. Selloni, "Electronic structure of defect states in hydroxylated and reduced rutile tio_2(110) surfaces," *Phys. Rev. Lett.*, vol. 97, p. 166803, Oct 2006.

[107] Z. Hu and H. Metiu, "Choice of u for dft+u calculations for titanium oxides," *The Journal of Physical Chemistry C*, vol. 115, no. 13, pp. 5841–5845, 2011.

[108] J. Deslippe, G. Samsonidze, D. A. Strubbe, M. Jain, M. L. Cohen, and S. G. Louie, "Berkeleygw: A massively parallel computer package for the calculation of the quasiparticle and optical properties of materials and nanostructures," *Computer Physics Communications*, vol. 183, no. 6, pp. 1269 – 1289, 2012.

[109] "A brief introduction to the abinit software package.," *Zeitschrift für Kristallographie - Crystalline Materials*, vol. 220, no. 5, pp. 558 – 562, 2017.

$\sqrt{17}$
$\Omega \int \delta e^{i\pi} = -1$
{2.7182818284}ϑφερτυθιοπσδφγηξκλ
χ^2
$\sum$
λ